DOCUMENT RESUME

ED 224 606 PS 013 283

AUTHOR Bond, James T.; And Others
TITLE Project Developmental Continuity Evaluation: Final Report. Volume I: Outcomes of the PDC Intervention.
INSTITUTION High/Scope Educational Research Foundation, Ypsilanti, Mich.
SPONS AGENCY Administration for Children, Youth, and Families (DHHS), Washington, D.C.
PUB DATE Sep 82
CONTRACT HEW-105-78-1307
NOTE 204p.; For related documents, see PS 013 284-285 and PS 013 337.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS Comparative Analysis; *Demonstration Programs; Early Childhood Education; *Federal Programs; Institutional Characteristics; Interpersonal Competence; Interviews; Longitudinal Studies; *Measures (Individuals); *Outcomes of Education; Parent Attitudes; *Program Effectiveness; Program Evaluation; Student Attitudes; Teacher Attitudes
IDENTIFIERS Developmental Continuity; Impact Studies; *Project Developmental Continuity; Project Head Start

ABSTRACT

The first of two volumes, this document reports an evaluation of Project Developmental Continuity (PDC), a Head Start demonstration project initiated in 1974 to develop program models which enhance children's social competence by fostering developmental continuity from preschool through the early elementary years. In general, the impact of program models on participating institutions, teachers, parents, and children is described. Following the first chapter's brief overview of PDC's objectives, development, and design, chapter 2 focuses on evaluation methods, discussing research questions, sample formation, measurement techniques, sample attrition, and data analyses. Chapter 3 specifies PDC's influence on local institutions. The first section of this chapter identifies the sources of data available for the evaluation, defines each guideline requirement for which implementation was assessed, and describes procedures used to aggregate ratings for purposes of characterizing implementation at each site. The second section of chapter 3 describes guideline-relevant features of PDC institutions, assessing guideline compliance and level of implementation. The three sections of chapter 4 describe, respectively, measurement of parent outcomes, the data analysis strategy, and findings from the analysis of parent outcome variables. In chapters 5 and 6, measurement of and results from the analysis of outcomes for teachers and children are discussed. Finally, a summary and explanation of the study's findings are provided in chapter 7. (MP)

PROJECT DEVELOPMENTAL CONTINUITY EVALUATION
FINAL REPORT

VOLUME I

OUTCOMES OF THE PDC INTERVENTION

James T. Bond

With contributions from:

John Berrueta-Clement
Sally Wacker
Barbara Bruemmer

Reviewed by:

José Rosario
Robert Halpern
Lynn Spencer

Technical assistance from:

Shirley Barnes
Deborah Edwards

September 1982

High/Scope Educational Research Foundation
Ypsilanti, Michigan

This report was prepared for the Administration for Children, Youth and Families, Office of Human Development Services, Department of Health and Human Services, under Contract No. HEW-105-78-1307, Dr. Esther Kresh, Project Officer. Views or conclusions contained herein should not be interpreted as reflecting the official opinion of the sponsoring agency.

ACKNOWLEDGMENTS

There are many whose contributions to the evaluation of Project Developmental Continuity should be acknowledged.

Very special thanks go to Esther Kresh, Administration for Children, Youth and Families (ACYF). As project officer for the national evaluation of Project Developmental Continuity since it began in 1974, she has provided continual support, encouragement, and direction. In addition, we wish to express our gratitude to other persons either now associated or once associated with ACYF and Project Developmental Continuity: Ray Collins, Jerri Klein, Austine Fowler, and Stephen Bedi.

It would have been impossible to collect evaluation data without the unflagging support and capable trouble-shooting of local PDC project coordinators: Robert Anastasi, Jesse Beard, Rowena Beck, Stephen Bedi, Tony Bosich, Mazarlo Carrillo, Glenda Dodd, Jerry Freddie, Faye Jerido, Deloris Johnson, Lynn Kagan, Beatrice Kenney, Sandy Kirby, Patricia Lanier, Mary (Wee) Leverman, Nancy Livingston, Betty Minor, Geraldine Sanders, Fannie Smith, Pat Tate, Patsy Thomas, and Cheryl Wilhoit. But perhaps our greatest appreciation should be extended to those many individuals who must remain nameless to protect their privacy--the children, parents, teachers, district and school administrators, Head Start staff and others who permitted what must have seemed endless interviewing, testing, and observation.

Collecting data for us was a dedicated team of local testers, interviewers, and observers who diligently tracked down children and parents, arranged data collection schedules with teachers, and rescheduled tests, interviews, and observations as necessary to get the information required.

We were fortunate to have an attentive Advisory Panel throughout the project who reviewed our work and shaped the direction of the evaluation: Eva Baker, Charles Billings, Jere Brophy, Robert Dixon, Robert Egbert, Ward Keesling, and Luis Laosa.

Last but not least are members of High Scope's staff who at some time in their careers helped with the work of the project. The following individuals were responsible for data processing: Barbara Bruemmer, Ann Hale, Robert Harvey, Helen Kiddon, Kim Marker, Jeffrey Moore, Kelly Naylor, Nancy Naylor, and Jane Oden. Mary Morris, assisted by Barbara Bruemmer, managed field data collection and training. Lynn Spencer provided administrative coordination and edited reports. Typing and proofreading were the responsibilities of Shirley Barnes, Deborah Edwards, Cathy Peterson, Gail Pheister, and Dana von Fange. Research assistants have included: Barbara Bruemmer, Judith Meece, Judy Platt, Mel Shelly, and Peter Williams. Research associates included: Ronald Duncan, Arthur Granville, Judy McNeil, and Allen Smith. David Weikart served as in-house monitor of the project for most of its life.

Finally, it is important that we acknowledge the very substantial contribution of John H. Love, who secured the evaluation contract and directed the project from 1974 to 1980. Marjorie Powell also spent a brief time as project director in 1980-81.

James T. Bond
Co-Director (1981-82)
PDC Evaluation

José Rosario
Co-Director (1981-82)
PDC Evaluation

Sally Wacker
Associate Director (1981-82)
Qualitative Analysis
PDC Evaluation

John R. Berrueta-Clement
Associate Director (1979-82)
Quantitative Analysis
PDC Evaluation

Table of Contents

Table of Contents
(continued)

Table of Contents
(continued)

Table of Contents
(continued)

Table of Contents
(continued)

PROJECT DEVELOPMENTAL CONTINUITY

The Administration for Children, Youth and Families (ACYF) initiated Project Developmental Continuity (PDC) in 1974 as a Head Start demonstration project. ACYF's goal was to stimulate the development of disseminable program models that would enhance children's "social competence" by fostering greater "developmental continuity" as they moved from Head Start into and through the early elementary years. To achieve this goal ACYF funded selected Head Start grantees across the country and provided technical support to Head Start centers and elementary schools to develop programs within a framework established by federal guidelines. Thus, ACYF set the general program objective and specified the basic parameters of the innovation process and of program operations, but relied upon local problem-solving to generate fully operational models.

THE PDC CONCEPT

In order to understand what PDC was about one must understand both its goal--"social competence"--and its means--"developmental continuity." These two concepts are rooted deep in ACYF's Head Start experience and, in one way or another, have always shaped the agency's work.

Social Competence

From its beginning in 1965, Head Start sought to provide for the needs of the "whole" child, not merely for intellectual or academic needs, which, however important, represent but one of many interdependent facets of the developing individual. Likewise Head Start objectives for child development have been holistic, stressing functional competence in the full range of life situations that children encounter as they grow up, rather than focusing single-mindedly on children's academic aptitude and school achievement. The phrase "social competence," describing this broad functional competence, became part of ACYF's official parlance more than a decade ago. Its early definition by a former commissioner of the Office of Child Development (now ACYF) is quoted here:

> *It may come as a surprise to the Nation that this preschool program was not mounted in hopes of dramatically raising IQ scores. ...Rather, the creators of Head Start hoped to bring about greater social competence in disadvantaged children. By social competence is meant an individual's everyday effectiveness in dealing with his environment...his ability to master appropriate formal concepts, to perform well in school, to stay out of trouble with the law, and to relate well to adults and other children.* (Edward Zigler quoted in Anderson and Messick, 1974)

ACYF's commitment to fostering the general social competency of children remains a driving force behind its policies and its actions.

Developmental Continuity

The concept of "developmental continuity" has shaped Head Start from its beginnings, and has in turn been shaped by the Head Start experience. The three continuity assumptions on which PDC was said to be based are central tenets of Head Start philosophy and programming (ACYF, 1977):

- *growth and learning occur as gradual and continuous processes;*
- *development is enhanced when programs are planned according to each child's needs, flow out of previous experience in and out of the home, and offer an orderly sequence of increasing complexity;*
- *the education of the child begins with the family and, therefore, the family's influence, stake and role in a child's development must be explicitly acknowledged in any early childhood education program.*

PDC differed from conventional Head Start programs by applying the concept of developmental continuity beyond the realm of Head Start in the context of public elementary schools.

PDC as an Extension of These Concepts

ACYF's decision to attempt the development of cooperative programs between Head Start centers and public elementary schools was an explicit response to problems identified during nearly a decade of Head Start experience, problems of "discontinuities" experienced by children moving from Head Start into and through the early elementary grades. The notion

that elementary schools were generally unresponsive to the needs of Head Start children, failing to support and build upon positive Head Start effects, was articulated during the first eighteen months of Head Start's existence (see, for example, Wolff & Stein, 1966). And Head Start proponents began calling for fundamental changes in the country's public elementary schools to ensure that Head Start's efforts on behalf of children from low income families would not be thwarted by subsequent experience in school.

In his State of the Union address to Congress in 1967, President Lyndon Johnson said, "We should strengthen the Head Start program, begin it for children 3 years old, and maintain its educational momentum by following through in the early years of school." (U.S. Congress, 1967) Six months later, a new Great Society program to "follow through" on Head Start was initiated. Known as Follow Through, this initiative has been managed and funded through the Office of Education (now Department of Education) from the fall of 1967 to the present, though it was evaluated only through the 1974-75 school year (U.S.O.E., 1977, Volumes I-V). The legislation creating Follow Through specified that the program would focus on children in kindergarten or elementary school who had been previously enrolled in Head Start or similar programs and that it would provide comprehensive services to children and encourage parent participation as described in the legislation authorizing Head Start. Thirty local projects were funded in the fall of 1967, and it was hoped that Follow Through would become a service program of the scale of Head Start by the following school year.

Follow Through's expansion into a nationwide social action program was not approved by a Congress growing leery of Great Society initiatives and worried by their costs. Thus, to justify its existence Follow Through was reconceptualized as a "planned variation experiment" in which alternative educational models for kindergarten through third grade would be developed and compared to determine their relative effectiveness in meeting the educational needs of children from low income families, especially children previously enrolled in Head Start. The task of developing alternative curricula was given to various research and development groups across the country who came to be known as "sponsors" of particular educational models. Sponsors were responsible not only for developing educational models but also for ensuring their implementation in the schools to which they were assigned at some number of sites. The total number of participating sites grew from 30 to a maximum of 178 at the height of the Follow Through project.

When Follow Through was redefined as a planned variation experiment comparing alternative elementary curricula, rather than a general service program extending Head Start provisions and philosophy into the public schools, there was an inevitable shift of emphasis at all levels from the broad objectives and methods of Head Start to the narrower concerns of traditional elementary education. The provision of comprehensive services

(mental and physical health, dental, nutritional, and social services) became the responsibility of local projects, and national efforts in this area were largely uncoordinated with other aspects of the Follow Through program. While parent involvement efforts of some sort were mandatory, sponsors varied greatly in the degree to which they developed educational models that systematically involved parents. Sponsors were also given considerable latitude in setting their program objectives, such that in the end few curriculum models explicitly addressed the needs of the "whole" child as these needs are defined by the Head Start notion of "social competence."

Follow Through's gradual drift away from the initial idea of extending Head Start into the elementary years was also a product of the decision to administer Follow Through through the Office of Education rather than the Office of Child Development (now ACYF), which was responsible for Head Start. However, in 1969, the Office of Child Development (OCD) took steps to strengthen the links between Head Start and Follow Through, at least in the area of educational curriculum, initiating their own planned variation at selected Follow Through sites. First eight, then eleven, Follow Through sponsors were invited to extend their educational models "downward" to selected Head Start centers feeding elementary schools implementing the sponsors' educational models. An additional sponsor with no Follow Through affiliation also got involved. That sponsor was OCD itself. Its "Enabler" model did not prescribe a particular instructional approach but provided each site with technical assistance to develop its own approach within the basic Head Start framework. (As will become apparent in the model discussion of the PDC program which follows, OCD's Enabler model foreshadowed the approach they would take in PDC.) Head Start Planned Variation (HSPV) projects were eventually mounted at 37 sites and served three cohorts of children between fall 1969 and spring 1972.

OCD (ACYF) hoped to link the evaluation results from HSPV with those from Follow Through to evaluate the combined program and to demonstrate the value of improving continuity between Head Start and school. Unfortunately, because the two programs and the two evaluations were managed independently, many HSPV graduates did not enroll in Follow Through schools and for those who did there was no control (comparison) group followed from HSPV entry through third grade. Consequently, HSPV and Follow Through were evaluated as separate interventions except for one very exploratory study (Weisberg & Haney, 1977) that found no evidence to support the hypothesized value of "educational continuity" beyond the kindergarten year. However, since the results of this study and of the national Follow Through evaluation were not available in 1973, they did not influence the design of PDC.

The decision to initiate PDC grew out of ACYF's persistent belief that the "momentum" generated by Head Start should be sustained by elementary schools, and ACYF's opinion that the concept of "follow through" had not really been put to a test. Thus, PDC, though shaped by previous experiments, did not replicate them:

- Formal institutional links were to be forged between Head Start centers and elementary schools to ensure coordinated program development and implementation.

- The project was to be managed and evaluated from start to finish by one agency of government, ACYF.

- The innovation would involve all classrooms at kindergarten through third grade levels in participating schools (rather than selected classrooms at these levels as generally happened in Follow Through).

- Specific curriculum decisions were to be left to local schools.

- The concept of developmental continuity was to be operationalized in all of its dimensions from Head Start through third grade by a process of local problem-solving within a general framework specified by ACYF (and derived from the basic Head Start model).

- Local projects were to be supported in developing and implementing their programs by outside technical assistants.

In these major respects PDC diverged very significantly from previous efforts.

ACYF identified four major categories of preschool/school discontinuity that PDC would explicitly address (ACYF, 1977):

Administrative elements: In Head Start and many preschools, class size is limited to 15 or 20 children with three adults, a teacher, an aide, and a volunteer. Scheduling can be flexible, according to children's needs and teachers' preferences. Today there are a number of program options in Head Start tailored to individual community and family needs. A center-based program may provide for variations in the number of days per week a child attends; some programs are home-based with or without regular attendance at a center.

Beginning with kindergarten, class size usually ranges from 24 to 40 children, with one teacher and possibly an aide. Usually the kindergarten is housed in a large building which serves many older children. Scheduling is less flexible because of the need to coordinate access to such facilities as playgrounds, libraries, resource centers, bathrooms, buses, etc.

Goals and objectives: Preschools tend to establish a foundation for the development of basic skills and to emphasize the socialization of the children. Some preschools do not have a well defined curriculum, and those which do may be at variance with that adopted by the state or local school system. Schools, on the other hand, tend to emphasize basic skills; teachers may be judged according to the academic progress of their pupils, as measured by national standards.

Structure of the program: Head Start was built on a comprehensive pattern of services which includes health and developmental support services as well as education. While many school systems are also concerned about comprehensive services such as health and nutrition, few have either the tradition or the necessary funds and resources to provide them. Head Start requires parent participation in classroom activities and in policy decisions. Over time, there has been little or no emphasis on parent involvement in school classrooms and decisions with respect to the lives of their children.

Expectations regarding children's behavior: Preschool children are usually expected to work cooperatively, to choose many of their own activities, and to talk with one another in order to foster language development. The teaching role is seen as that of facilitating. In school, children are usually expected to work independently, and cooperative activities are less often available. Most of the time, they are expected to stay in their seats or in designated classroom areas, and talking may be discouraged except for limited purposes related to instruction. The teacher's role is seen as directing.

The challenge to PDC program developers was to reduce these specific discontinuities by developing and implementing new structures and processes within and between Head Start centers and elementary schools.

In large measure, all children experience similar discontinuities upon entering school, and many suffer no obvious ill effects. That is, they gradually "accommodate" to the rather strange environment of school and function more or less effectively within it. The assumption underlying PDC, however, was that some children, particularly children like those served by Head Start, are often unable to accommodate effectively to conventional kindergarten and early elementary environments because of extreme environmental discontinuities between the cultures of home and school, and between Head Start and school. Moreover, it was thought that individual developmental discontinuities often arose because conventional school programs did not attend to unmet basic needs of the child by providing comprehensive services and did not systematically foster coordinated development of all aspects of social competence. As with Follow Through, PDC's strategy for improving continuity emphasized changing schools to accommodate Head Start children rather than changing Head Start children so that they might better accommodate to school. The only area in which there was an expectation that schools might significantly influence Head Start was academic curriculum development, since Head Start and elementary school staff were supposed to collaborate in developing a continuous educational program from preschool through at least third grade.

THE PDC PROGRAM

PDC was both a national and a local intervention. Although ACYF applied a single intervention model across all sites, each site developed its own intervention within the ACYF framework, producing substantial variation in the "treatments" experienced by teachers, parents, and children in different communities. In this respect, PDC was very much like Head Start--a national program giving considerable autonomy to local centers within limits set by common performance standards.

ACYF's approach in PDC was to define a framework for innovation at all sites, but to encourage local problem-solving and locally generated variations within this framework. ACYF's framework was operational mainly at the institutional level, specifying certain policies, procedures, and organizational structures within and across Head Start centers and schools that had to be employed in developing and implementing local PDC programs. The specific "treatments" for teachers, parents, and ultimately children were not operationally defined, though the objectives of these treatments were described. It was assumed that the same objectives might be achieved in rather different ways by different sites.

What might be called the PDC "product concept" was articulated in the following objectives (ACYF, May 1977):

> *Implement a sequenced and continuous educational program for children as they move from the Head Start through the primary grades. This program will emphasize social competence, teaching of basic skills, and individualized instruction. It must include developmental activities which encourage the physical, intellectual and social-emotional growth of children.*
>
> *Provide ongoing training for Head Start and school parents and staff and Council members in the areas of (a) child growth and development and (b) Project Developmental Continuity's philosophy, goals and objectives. Training must be oriented to helping staff meet the developmental needs of the total child.*
>
> *Involve parents in the Head Start and school experiences of their children, promote parents' understanding of the continuity of the child's development and the importance of continuity of experiences, and enhance parent participation in the decision-making process.*
>
> *Involve teachers and parents in meeting the child's social, emotional and intellectual needs in ways appropriate to his developmental level.*

Promote continuity in the nutritional, medical, dental and social services provided to children as they move from a Head Start program to the primary grades.

Insure that handicapped children receive individualized services within the context of the regular Head Start and elementary school program.

Insure that children from different language and cultural backgrounds receive individualized services within the context of the regular Head Start and elementary school program.

Develop necessary administrative mechanisms that will guide the planning and implementation of Project Developmental Continuity projects.

Explore various strategies for achieving continuity by allowing for variations in the programs at the local level.

These objectives were further operationalized in a set of guidelines, which laid out "product specifications" for which local programs were held accountable.

PDC Guidelines

The PDC Guidelines provided ACYF with a modicum of control over the development and implementation of programs in local projects. Moreover, they provided local projects with a ready-made framework for innovation that might have been difficult for some to formulate independently guided only by general statements of program development objectives. Guidelines for a Planning Year were distributed in 1974 to all agencies invited to submit proposals for program development projects. These initial Guidelines were subsequently revised and refined as necessary, this process culminating in a final set of Implementation Year Guidelines issued in 1975 when the program development effort was underway. (Subsequent references to "the guidelines" pertain to those issued in 1975.) The Guidelines were supplemented by "Program Letters" to the projects which clarified issues as needed.

The PDC Guidelines were organized into seven "component areas." Within each area, "basic principles" were stated and "required elements" specified (ACYF, 1975):

1. *Basic Principles are general statements of philosophy pertaining to the Component Areas which must be addressed by all sites when designing activities in those areas.*

2. *Required Elements are more specific program activities or details which must be included within the Component Area during the operational year.*

ACYF encouraged natural variation in response to local situations:

> *Sites may design locally appropriate methods or activities within each component area, provided that the basic principles are addressed and the required elements are included. Regardless of the strategies decided upon for full component coverage, the total plan must be suitable to the particular needs of the locale, and must be satisfactory to the community. Local ethnic, cultural and language characteristics must be taken into account.*

ACYF's posture toward local program variation reflected both Head Start policy regarding community control and an understanding, based on long experience, that without a sense of "local ownership" innovations are unlikely to take root.

Guideline requirements in each of the seven component areas are described briefly here, and at greater length in Chapter II which is concerned with specifying the PDC treatment at each site:

- Administration. Each project was to hire a full-time PDC coordinator as well as full- or part-time support services and parent involvement coordinators. Each site was also to create a PDC Council composed of representatives from the following groups: parents of PDC Head Start or elementary school children; members of the Head Start Policy Council and local school board; Head Start and elementary school administrators; Head Start and elementary school staff; and local community groups. This Council was to be responsible for the overall operation of the PDC project.

- Education. Concern for the "whole" child was emphasized in all the education Guidelines. Sites were required to adapt or adopt a compatible, coordinated curriculum that provides experiences for children appropriate to their developmental levels, interests, and needs. The use of individualized instruction and diagnostic-evaluative systems would facilitate teacher awareness of the uniqueness of each child along with the physical, intellectual and socio-emotional growth of each child.

- Services for bilingual-bicultural and/or multicultural children. Guidelines stressed the importance of taking into account the different ethnic and cultural backgrounds of children. Classroom activities and materials were to reinforce children's pride in and understanding of their

background and provide opportunities for children to learn about and appreciate the cultures of others. Teachers were to be made sensitive to the needs of multicultural children and to involve parents in their children's educational program.

- Bilingual-bicultural demonstration projects. A special set of basic principles and required elements was written for those sites designated as bilingual-bicultural demonstration projects. These Guidelines stated that all components at these sites were to incorporate a bilingual-bicultural approach. An educational and social setting was to be provided that was based on the child's primary language and culture. The bilingual-bicultural educational approach was to build upon strengths the child brought to the learning situation, to expand upon the child's native language and to make use of the child's native language for instructional purposes.

- Services for handicapped children. PDC was committed to the concept of mainstreaming. The Guidelines further required a yearly survey of handicapped children, procedures for early diagnosis and evaluation, special resource teachers, and special training for classroom teachers in working with handicapped children.

- Parent involvement. Concern with involvement of parents in school activities permeated the Guidelines. Sites were required to develop coordinated parent programs that involved parents in all phases of program planning, operation and evaluation. Guidelines also required that programs try to involve parents in classrooms, in the Council, in component subcommittees, in training sessions or workshops, and in planning PDC activities.

- Developmental support services. Guidelines for this component defined the kinds of services that had to be made available to all PDC children. The nutritional, medical, dental, mental health, and social services needs of children were to be assessed upon entry into the program and arrangements made to provide needed services. There was to be a consistent and complete record-keeping system, contact with community resources, and information provided to parents about their children's needs and the availability of community resources.

- Training. Guidelines stressed the need for ongoing training activities and called for a schedule that included sessions related to each of the component areas with agendas that targeted diverse audiences. For example, the Guidelines called for training parent volunteers to work in the classroom, training for teachers to sensitize them to the special needs of multicultural children, training for PDC Council members in policy- and decision-making skills, and training for teachers and administrators in how to work with parents. (For purposes of implementation rating these requirements were merged into other component areas with which they were functionally associated.)

Though never explicitly acknowledged by ACYF, it was readily apparent to all who read ACYF's project objectives and requirements that typical public elementary schools would have to do a lot more changing than typical Head Start centers if PDC Guidelines were to be fully implemented.

Take, for example, the proposed "PDC Council" (Administration component). Head Start staff are accustomed to working not only with but under the direction of Head Start Councils, whose membership includes parents and community representatives. Public school staff, however, are generally quite unaccustomed to such arrangements, and school district administrative systems are not designed to accommodate formal advisory inputs from such Councils, much less their management of school programs. While the requirement that public schools be concerned with the "whole" child (Education component) certainly did not contradict educational philosophy and values at the early elementary level, the expectation that this concern would be "active" and manifested in an integrated educational program entailed substantive change in elementary schools. And though requirements for participation by parents (Parent Involvement component) had some precedent in regulations governing Title I and other federally funded programs already operating in most participating schools, PDC Guidelines generally required more central involvement of parents in school decision-making than did other federal programs, even going so far as to suggest that parents might influence curriculum within the classroom. Head Start programs had long offered parents opportunities to participate both in educational settings and in the creation of these settings, but the creation of educational settings in elementary schools had always been the prerogative of professional educators. It remained to be seen whether or not teachers and administrators would be open to significant parental inputs affecting educational programming. In these and many other ways, PDC Guidelines required public elementary schools to change in fundamental ways, ways that promised to alter the schools far more than Head Start, which after all had inspired the PDC Guidelines.

Site Selection

The site selection process began in 1974 and involved several steps. Each regional ACYF office[1] and the Indian and Migrant Program Division asked various Head Start grantees within each region to complete a questionnaire to determine the feasibility of implementing PDC at their sites. ACYF's ultimate judgment of site suitability would take into account a broad range of factors. First, ACYF was concerned that there be evidence of local interest in the goals and methods of PDC and of willingness to operate within the framework of the Guidelines. Second, ACYF wanted

[1]The ACYF regional offices are located in Boston (Region I), New York City (Region II), Philadelphia (Region III), Atlanta (Region IV), Chicago (Region V), Dallas (Region VI), Kansas City (Region VII), Denver (Region VIII), San Francisco (Region IX), and Seattle (Region X).

assurances that the conditions of the planned PDC evaluation could be met: Was there a possibility of forming a local comparison group comprising Head Start centers, schools, teachers, parents, and children that would be roughly "equivalent" to the PDC sample and willing to participate in the evaluation? And were there enough potential PDC and Comparison children to permit a longitudinal evaluation (Head Start through third grade) in the face of inevitable attrition? Third, ACYF wished to select sites that represented two types of physical relationship between Head Start and the public schools. One type was referred to as the "Early Childhood School" (ECS) because Head Start was located within the elementary school building. The other type was called "Preschool-School Linkage" (PSL) because Head Start and elementary programs were housed in separate physical plants. Based on responses to the initial survey, officials in each region identified four sites as most suitable candidates.

An ACYF review panel then invited the two most suitable sites from each Region to submit formal proposals. Staff from the national and regional offices reviewed these proposals and visited each site to meet with grantee staff and with Head Start and public school teachers, parents, and administrators. Since PDC was intended to extend beyond Head Start into the elementary grades, U.S. Office of Education and state education agency personnel were also involved in the final review process. Through this process two sites were selected from each of three Regions (II, III, and VIII), and one site, from each of the remaining Regions. Three of these 13 sites were designated Bilingual-Bicultural Demonstration Projects serving Hispanic children. Two additional sites were selected to represent the Indian and Migrant Program Division of ACYF, raising the total number of local projects to 15.

Operation of the program began in the fall of 1974 at 14 sites and in January 1975 at the fifteenth. The entire first year of program operation was devoted to program planning. Staff were hired, component area task forces were appointed, and detailed plans for actual implementation were gradually developed.

During Year II, 1975-76, 14 sites (one had withdrawn), comprising a total of 42 Head Start centers and elementary schools, began to implement their plans on a trial basis. Year III, 1976-77, was designated the official "Implementation Year"--the year in which all local programs were expected to be fully operational, and the year in which the cohort of children that would constitute the evaluation sample would be enrolled in Head Start. At the end of Year II, a second site dropped out, leaving 13 local projects to implement their developmental continuity programs. At the end of Year III, ACYF committed itself to funding the remaining projects through the end of the 1980-81 school year when the evaluation cohort of children would graduate from third grade and PDC.

Subsequently, one more site dropped out of the project. Another was excluded from the evaluation. And a third was included in the evaluation but is excluded from analyses conducted for this report because no local comparison group could be found to provide a point of reference in estimating PDC program effects. Thus, the longitudinal evaluation of PDC is restricted to data from 10 of the original 15 sites.

Technical Assistance

By providing local projects with continuing external technical assistance, ACYF acknowledged and attempted to overcome common limitations of the local problem-solving approach to program development. Without exceptional leadership and management, problem-solving efforts with the scope of PDC are likely to lose direction, bog-down, and/or expend enormous energies "reinventing the wheel." Technical assistance was intended to keep the program development process moving, on track in relation to objectives and Guidelines, and well-informed.

Technical assistance to local projects was provided first by Huron Institute of Cambridge, Massachusetts, then by Pacific Consultants of Washington, D.C. During Years I and II, each site was visited several times by Huron Institute's field staff whose role was to facilitate the local problem-solving process. Of particular concern were ensuring participation in the planning process by all individuals and groups specified in the Guidelines, attention and conformance to Guidelines in all seven component areas, and consideration of a broad range of programmatic alternatives before final decisions were made. Huron Institute staff also assisted ACYF in planning and conducting national workshops that provided a continuing forum for discussion of implementation, funding, and evaluation issues, as well as opportunities for formal and informal exchange of ideas among project staff from across the country.

Pacific Consultants assumed responsibility for PDC technical assistance in Year III of the project and continued to provide various services to local projects and ACYF until federal support of the demonstration terminated in June 1981. During this period, Pacific Consultants staff made periodic site visits to help local project staff address issues of implementation and continuing program development. In addition, they helped ACYF plan national conferences and they published occasional PDC newsletters intended to foster better communication among local projects and to represent PDC to the larger community.

Compliance Monitoring

ACYF entered into agreements with local projects with the understanding that their contracts would be renewed only if they both adhered to PDC Guidelines and satisfied the conditions of the evaluation. ACYF soon relaxed its requirements that the conditions of the evaluation be fully met, and they continued funding to two sites that did not satisfy evaluation criteria. However, ACYF continued to monitor conformance to the Guidelines.

During the first year of the project, compliance monitoring was the responsibility of Development Associates of Arlington, Virginia under subcontract to High/Scope Educational Research Foundation, the evaluator.

This subcontractual arrangement was intended to dissociate the inherently threatening and controversial monitoring function from project evaluation, the results of which would not directly affect continued funding of local projects and the process of which required the close cooperation of local PDC staff. In fact, the evaluation and monitoring functions overlapped substantially, given the evaluation's initial focus on program implementation. Therefore, third party monitoring of local projects was eliminated as a separate and distinct task in Year II. From that point forward, ACYF assumed full responsibility directly and through its regional offices, for determining whether or not local projects had satisfied the terms of their contracts. Site implementation reports by the evaluator were considered in this review process as one source of information.

Most sites found it relatively easy to satisfy the "letter" of the Guidelines; however, as High/Scope's report on program implementation through Year III indicated, there was considerable cross-site variation in the degree to which the "spirit" of the Guidelines had been implemented (Smith, Love, Morris, Spencer, Ispa, & Rosario, 1977). Though ACYF did not terminate any site for noncompliance, two sites eventually elected to withdraw from the project rather than comply fully with the Guidelines.

Relative Scale and Cost of PDC

In terms of number of sites served, PDC was substantially smaller than either Head Start Planned Variation (maximum, 37 sites) or Follow Through (maximum, 178 sites). In terms of number of children served per year, PDC (with 6,000 to 7,000 children enrolled per year from 1976 on) was somewhat larger than HSPV (with a maximum of 4,974 children enrolled in 1970-71 [Smith, 1973, p. 46]) and much smaller than Follow Through (with total enrollments ranging from 50,000 to 100,000 children from 1969-1975 [U.S.O.E./Haney, Vol. V, 1977, p. 21]).

Estimating the relative costs of PDC versus other demonstration programs is more difficult than one might imagine, and any thorough treatment is beyond the scope of this report.[1] However, one simple and potentially informative comparison can be made between PDC and Follow Through. During the 1967-77 program year PDC grants to local projects amounted to $186 per child, while Follow Through grants to local projects during the preceding year (1975-76 amounted to $572 per child (with considerable regional variation in both figures). These amounts exclude modest technical

[1]The interested reader is referred to the following reports for cost data pertaining to PDC, Follow Through, and HSPV, respectively: Development Associates and High/Scope Educational Research Foundation (1977); U.S.O.E./RMC (1977); McMeekin, R. W., (1973).

assistance costs for PDC (less than $20 per child) and very sizable costs associated with outside "sponsorship" of educational programs in Follow Through ($152 per child). Disregarding inflation from one year to the next and including or excluding technical assistance/sponsorship costs, it can be seen that the direct costs to the federal government were much smaller on a per capita basis for PDC than for Follow Through, reflecting PDC's very different change strategy which relied upon local problem-solving to produce innovations that might be sustained without perpetual external support.

The incremental costs of PDC--i.e., the additional cost per child of implementing PDC versus whatever programs were otherwise being implemented in non-PDC schools--were estimated in a cost study conducted during the 1976-77 program year (Development Associates & High/Scope Foundation, 1977). Interestingly, the final estimate of incremental cost, based on the "value of resources utilized" per child including in-kind contributions and volunteer hours, was very nearly equal to the PDC grant per child--$192 versus $186 across the nine sites in the study. Because the economic value of volunteer hours and all manner of in-kind contributions were confounded with actual cash outlays, this figure may well over-estimate the "out of pocket" cost of implementing PDC.

Related to the question of PDC's incremental costs, we have observed that the politics of local educational systems work against sustained differential treatment of schools within the same community, particularly differential allocation of resources. Anecdotal data collected during site visits from 1976 through 1981 suggest that local education agencies frequently respond to new infusions of federal funds by reallocating resources to maintain a rough parity among schools in their district (controlling for characteristics of the population served). Thus, while inputs to PDC and non-PDC schools within the same community may have differed in kind, it seems much less likely that they differed appreciably in amount over the life of the project. One striking example of this reallocation phenomenon will be mentioned to illustrate the point. At the site where the incremental cost of the PDC program was found to be largest in 1976-77 ($603 per child), the school district is known to have begun pooling all federal monies (apparently with at least tacit approval from the relevant agencies) in order to allocate them more equitably across all schools, in keeping with the objectives of the program of origin, but ignoring program boundaries and many specific regulations. Thus, any early difference in costs per child in PDC versus non-PDC schools was quickly attenuated, if not completely eliminated, by creative management of federal funds.

PDC's Cessation

ACYF's funding for local PDC projects terminated in June 1981 after one cohort of children had graduated from third grade having completed the five-year developmental continuity program. It was hoped that at least some features of PDC would become institutionalized--i.e., be adopted and sustained by local institutions at their own expense (if any). The decision not to make further investments either for purposes of developing training materials related to local programs or for dissemination of locally developed models to other communities presumably reflects ACYF's judgment that such an investment would not have been cost-effective.

II

THE PDC EVALUATION

ACYF requested an evaluation of PDC that would examine both the *process* of program implementation, and *impacts* of this process on institutions, parents, teachers, and children. The evaluation was to proceed in two phases. The first, extending from 1974 through 1978 (Years I-IV), was aimed at evaluating the initial process of PDC and at determining the feasibility of conducting a five-year longitudinal study that would follow one cohort of children from the time they entered Head Start through their third grade year. The second phase, extending from the fall of 1978 through the spring of 1982, was intended to evaluate the impacts of PDC longitudinally, while continuing to document major aspects of the program process in order to explain impacts. Having concluded, on the basis of findings from phase 1, that it was feasible to conduct a longitudinal study of PDC, ACYF proceeded with the second phase as originally planned. The High/Scope Educational Research Foundation has conducted both phases of the evaluation. The results of the first phase have already been published (Love, Granville, & Smith, 1978; Smith et al., 1977) and are incorporated here only as necessary for the purposes of the longitudinal study.[1]

DERIVATION OF RESEARCH QUESTIONS

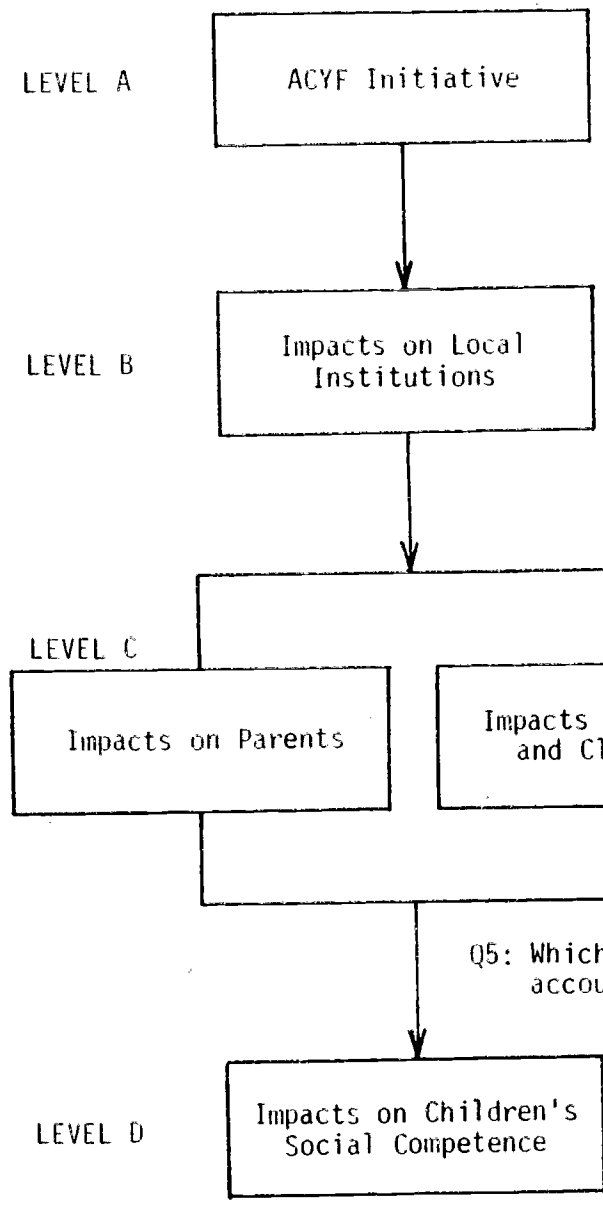
Research questions were shaped by the *intervention model* implicit in PDC. Figure II-1 illustrates this model in simplified form, together with the major research questions addressed by the longitudinal evaluation. It is fundamentally a "top-down" model in which "bottom-up" responses over time were expected to stimulate and guide the change process, and in which interactions across all levels and among all components were expected to sustain the innovation. The solid unidirectional arrows linking Levels A through D portray the assumed temporally ordered sequence of effects across levels, leading from ACYF's initiative to child outcomes. Feedback loops from Level D (child outcomes) to Level C (parent and teacher outcomes) to Level B (institutional outcomes) are acknowledged but not represented.

The research questions listed in Figure II-1 represent those questions of highest technical priority in the longitudinal evaluation. There are, of course, larger and more significant questions of social policy that the evaluation should ultimately address inasmuch as it can. Some of these questions were clearly implied, if not explicitly articulated, by ACYF when

[1]The summary chapter of the "Final Report of the PDC Feasibility Study, 1974-1977" (Love, Granville, & Smith, 1978) is reproduced in Appendix A.

Figure II-1

The PDC Intervention Model and Associated Research Questions



Q1a: Did Head Start centers and public elementary schools participating in PDC implement the institutional features prescribed in the PDC Guidelines?

Q1b: Did implementation of PDC Guidelines produce institutional features in PDC centers and schools that were different from those present in comparable non-PDC centers and schools in the same community?

Q2: Did PDC parents exhibit attitudes, knowledge, and/or behaviors toward school and toward their own children that were different from those of non-PDC parents and, at the same time, compatible with the intentions of the Guidelines?

Q3: Did PDC teachers exhibit attitudes, knowledge, and/or behaviors toward children that were different from those of non-PDC teachers and, at the same time, compatible with the intentions of the Guidelines?

Q5: Which program effects on parents and teachers accounted for impacts on children?

Q4: Did PDC children exhibit greater social competence than would have been expected had they not participated in the program?

it initiated PDC. For example: *What strategies should ACYF pursue to more effectively support the transition of children from Head Start to elementary school? Is an educational change strategy or program model development strategy that relies heavily upon local problem-solving viable?* Other questions with important policy implications emerged over the course of the study. For example: *Can programs seeking to reduce the environmental discontinuities that children experience in the early elementary grades be effective given the high residential mobility of families in contemporary society? On what basis should decisions to conduct major longitudinal evaluations of educational programs be made?* Though this evaluation cannot provide definitive answers to such questions, it is hoped that its findings and lessons learned from the experience will influence the policy-making process. Policy implications of the evaluation will be considered in the concluding chapters of this volume and Volume II. The immediate focus of this report, however, is the set of questions posed in Figure II-1.

Model Specification and Expectations

As a cross-site intervention the PDC treatment was operationally defined at Level B (local institutions) as that particular configuration of institutional features prescribed in the PDC Guidelines. ACYF's intervention model assumed that the desired child outcomes (Level D) would be produced if this institutional configuration were implemented--i.e., *if B, then D.*

Impacts at Level C (parents and teachers) were largely unspecified by the intervention model in operational (i.e., measurable) terms. Though it was certainly assumed that the behaviors of parents and teachers toward children would be primary mediating variables between institutional impacts and child outcomes, ACYF deliberately avoided identifying specific parent/teacher-child interactions as "required" parts of the PDC treatment. Instead, ACYF assumed that parents and teachers at each site, and perhaps each parent and teacher, would develop, within the PDC institutional context, a pattern of interacting with children that was "developmentally appropriate," "individualized," "culturally appropriate," and "effective" in that it enhanced each child's social competence. It was at this level, then, that ACYF encouraged "unplanned" or "natural" variations in the PDC treatment.

Thus, *as a local, within-site intervention* the PDC treatment was defined in terms of institutional features prescribed in the Guidelines *plus* local variations in institutional features, teacher behaviors, and parent behaviors that were in some measure unique to each site. When Level C (parents and teachers) is considered along with Levels B (institutions) and D (child outcomes), the effects formula reads: *If B, then D by means of impacts at C not specified a priori.*

Impacts at Level D (child outcomes) were simply defined as "social competence" in all of its aspects. Operationalization of the social competence construct was undertaken by the evaluator in collaboration with ACYF staff and a National Advisory Panel.

Central Research Questions

The evaluation of PDC has attempted to assess the validity of the effects formula implied by the intervention model:

> *If the institutional configurations prescribed in the PDC Guidelines are implemented (Level B), then desired child outcomes will be realized (Level D) by means of unspecified impacts on parents and teachers (Level C).*

Research questions were formulated accordingly.

Since the success of the PDC intervention was assumed to hinge upon first achieving institutional impacts (Level B), the question of whether these impacts were achieved will be addressed in the next chapter (III) of this volume, before considering outcomes at Levels C and D:

> Q1a: Did Head Start centers and public elementary schools participating in PDC implement the institutional features prescribed in the PDC Guidelines?

However, the expectation "if B, then D" was based upon certain assumptions about non-PDC centers and schools that may not have been true. Specifically, it may not have been necessary to change existing institutions in order to comply with particular PDC Guidelines. Other federally funded programs operating in PDC communities also required parent participation of some sort. State and/or federal laws already required individual educational programs for handicapped students. And virtually every school district in the country, in the evaluator's experience, was committed to individualizing instruction in one form or another. Thus, it was conceivable that PDC required relatively few changes in existing institutions, changes involving primarily formal administrative linkages between Head Start and public elementary schools. Thus, question 1b was posed:

> Q1b: Did implementation of PDC Guidelines produce institutional features in PDC centers and schools that were different from those present in non-PDC centers and schools within the same community?

Unless systematic differences were found between PDC and non-PDC institutions, there would be no reason to expect differences in child outcomes, or in parent and teacher outcomes, due to PDC versus non-PDC treatments. Answers to questions 1a and b in Chapter III represent a specification of the PDC treatment at the institutional level and establish expectations regarding program effects at Levels C (parents and teachers) and D (child outcomes).

Although the intervention model did not prescribe specific links between institutional impacts and child outcomes, it presumed the existence of such links and implied their general shape. PDC parents and teachers were expected to be changed by their participation in new institutional structures. It was assumed that they would in turn alter classroom and home environments directly affecting children in ways that would enhance children's social competence. Chapters IV and V of this volume examine parent and teacher/classroom outcomes, respectively, in an effort to uncover program-related differences in the proximal environments of PDC and non-PDC children:

Q2: Did PDC parents exhibit attitudes, knowledge, and behaviors toward school and toward their own children that were different from those of non-PDC parents and, at the same time, compatible with the intention of the PDC Guidelines?

Q3: Did PDC teachers exhibit attitudes, knowledge, and behaviors toward children that were different from those of non-PDC parents and, at the same time, compatible with the intentions of the PDC Guidelines?

Lacking clear a priori specification of effects at this Level (C), it was necessary to cast a broad measurement net in hope of capturing any important general and site-specific program effects that might occur.

Whether the PDC intervention succeeded in enhancing children's social competence (Level D) is considered in Chapter VI:

Q4: Did PDC children exhibit greater social competence than would have been expected had they not participated in the program?

The question of how PDC achieved its impacts on children by way of impacts on parents and teachers will also be addressed in Chapter VI as appropriate given child outcome findings:

Q5: Which program effects on teachers and parents account for impacts on children?

It was ACYF's hope that a posteriori identification of critical parent and teacher outcomes at Level C might contribute to the further development of the PDC intervention model and of agency policy.

The questions posed above will be further elaborated and given full operational definition as they are addressed in the chapters that follow. Additional research questions are addressed in Volume II of this report, The Process of Program Implementation in PDC, which examines dynamic features of the process whereby ACYF's initiative interacted with prior characteristics of local institutions and community context to produce the particular patterns of institutional impacts described in Chapter III of this volume.

EVALUATION DESIGN

As already mentioned, the study design incorporated a "control" or comparison sample at each site as a point of reference in estimating PDC effects--i.e., in estimating what PDC centers, schools, parents, teachers, and children would have been like without ACYF's intervention. This comparison sample is hereafter referred to as the Comparison group (abbreviated, COMP or C). At each site, the PDC and Comparison groups together comprised one cohort of children who entered Head Start in the fall of 1976 and were expected to graduate from third grade (the end of the PDC program) in the spring of 1981. In the longitudinal study reported here, data were collected only for children in this evaluation cohort, their parents, their teachers, and their centers and schools. *Since PDC and Comparison children had comparable Head Start experiences, any differences in later outcomes were expected to result from differences in their experiences at transition to elementary school and during the early elementary grades.*

PDC as a Quasiexperiment

Since centers, schools, parents, teachers, and children were not randomly assigned to PDC and the Comparison group, the evaluation design was quasiexperimental rather than truly experimental. Though great pains were taken to select a Comparison group that was equivalent in all respects to the PDC sample prior to the PDC intervention, the selection procedures employed did not guarantee initial equivalence. And even though the PDC and Comparison groups may have been equivalent in all respects at program entry, it is entirely possible that attrition from the sample over the course of the five-year longitudinal study may have been nonrandom, making the longitudinal PDC and Comparison samples nonequivalent with respect to entry-level characteristics. Threats to the internal validity of the evaluation design posed by nonrandom formation of PDC and Comparison samples and by possibly nonrandom attrition cannot be overemphasized, for the evaluation was specifically designed to use measures of Comparison group institutions, teachers, parents, and children as unbiased estimates of what the PDC sample would have been like but for PDC. Issues of PDC/Comparison comparability and the selection of analytic strategies to cope with potential noncomparability are discussed later in this section.

PDC, One or Many Interventions?

PDC was both a single, cross-site intervention and many, within-site interventions. ACYF took very similar initiatives with respect to all sites, which is to say, there was a single intervention by an agency of federal government. However, this single intervention set into motion innovation processes at each site that created different local interventions tailored to different local contexts.

The fact that PDC was both one and many interventions had important implications for the evaluation design. It was not only conceivable, but to be expected, that some local intervention models would be more effective than others in realizing the "spirit" of PDC's Guidelines and objectives. Moreover, it was possible that some local interventions would be more effective than others in realizing PDC's ultimate goal of enhancing children's social competence. Consequently, *it was necessary to address research questions at the site level before they could be meaningfully addressed at the cross-site level.*

The remainder of this chapter discusses sample formation, measurement, sample attrition, and data analysis. Particular attention is given to the development of a data analytic strategy that would permit informative site-level evaluations in the face of many limitations imposed by the quasi-experimental design.

Sample Formation

The PDC centers, schools, teachers, parents, and children to be included in the longitudinal evaluation were determined by initial agreements with each site designed to ensure that there would be enough children in the evaluation cohort to permit meaningful longitudinal follow up. At each site, the evaluation sample comprised:

- Children attending one or more Head Start centers participating in the PDC program who subsequently entered public elementary schools also participating in the PDC program;
- Parents of these children;
- Teachers and other staff working in the centers and schools these children attended; and
- Participating Head Start centers and schools.

At the ten sites included in the longitudinal study, the number of PDC Head Start centers ranged from 1 to 2 per site; the number of PDC elementary schools, from 1 to 3.

The originally proposed evaluation design entailed forming a non-PDC sample at each site that closely resembled the PDC sample with respect to pre-PDC characteristics and characteristics over which PDC could have no influence. During the 1975-76 school year, a pilot study was conducted to assess the feasibility of forming such Comparison samples at each site (Granville, McNeil, Meece, Wacker, Morris, Shelly, & Love, 1976). First non-PDC elementary schools serving relatively large numbers of Head Start graduates were identified in each community. These schools and their associated Head Start centers then formed a pool from which Comparison centers and schools might be drawn. In two communities with centralized Head Start centers, there was no possibility of finding centers or Head Start teachers unaffiliated with PDC. Limited demographic data and rather extensive test data collected on PDC and pilot Comparison children during 1975-76 suggested that it would be possible to form roughly equivalent samples of Comparison children at each of the ten sites to which the longitudinal study has been restricted. Moreover, information collected about potential Comparison schools indicated that one or more of these schools in each community was quite similar to the PDC school or schools in terms of staffing, availability of special programs, physical facilities, student population size, and characteristics of students served. The most similar Comparison schools were termed "matched" Comparison schools, and it was expected that subsequent collection of Comparison teacher and institutional data would be concentrated there.

When originally planned, the evaluation design presumed that PDC and Comparison sample sizes would not fall below 30 children/families per group per site. Given this number of children and families per site, it would be possible to conduct site-level analyses with some reasonable hope of identifying moderate but reliable program effects if any existed. The Year II pilot study (Granville et al., 1976), however, raised serious doubts about the adequacy of longitudinal sample sizes. Based on data collected from local Head Start centers and schools during 1974-75 and 1975-76, various projections were made of the probable proportions of Head Start children who would remain through third grade in elementary schools they entered at kindergarten. The final best estimate was 35-40%, which meant that, given average Head Start enrollments, only 21-24 children per group per site on the average would complete third grade in the schools they entered at kindergarten. In spite of these findings, a decision was made to proceed with the longitudinal evaluation as planned with the hope that attrition would be less than projected or that the analytic strategy could be modified to cope with smaller than anticipated site-level samples.

Measurement

The decade preceding PDC was replete with major evaluations of federally funded educational interventions. Few of these evaluations, and none of the larger ones, had been particularly successful--leaving, as they did, enormous doubts regarding the effectiveness of the programs evaluated, providing little or no information about why particular programs seemed to work or not to work, failing to measure the actual objectives of programs, and failing even to specify the treatments evaluated other than to name them.

When ACYF conceived of Project Developmental Continuity, they were adamant that the evaluation of PDC represent at least an improvement upon earlier evaluations. Thus, ACYF insisted upon attention to the process of program implementation from the very beginning of the evaluation so that the treatment could be specified and outcomes might be explained. ACYF also insisted that every effort be made to evaluate the ultimate objectives of PDC--enhanced social competence of children--rather than employ IQ and achievement tests as primary outcome measures simply because of their convenience. And ACYF required that two batteries be formed for assessing child outcomes--one to measure common objectives across all sites and one to measure objectives that were relevant only to particular local programs. As already noted, ACYF planned to continue the evaluation into its longitudinal phase only if satisfied that the design offered some reasonable hope of identifying relevant impacts if they occurred.

In response to ACYF's request and out of shared commitment to ACYF's goals, the evaluator invested a considerable portion of available resources in the development of a suitable measurement battery for PDC. With advice from ACYF and a national Advisory Panel, existing instruments were reviewed and tested, and new instruments were developed, in a process that began in 1974 and did not terminate until the final data collection in the spring of 1981. When all is said and done, we must conclude that this evaluation has not overcome all of the measurement limitations of previous work, but it is our hope that these efforts to more thoroughly measure both the process and the product of the intervention will provide greater insight into what went right or wrong, and why.

A particularly sensitive area of measurement was children's "social competence"--everyday effectiveness in dealing with the environments of home, school, and community. It is an attractive construct to educational researchers, and to most people; however, it is extraordinarily difficult to operationalize, to define in measurable terms. This evaluation does not presume to have adequately measured children's social competence. Though we may have taken a small step in the proper direction, we readily admit that we have fallen short of the target. It will remain for later evaluations to fully realize the goal. Toward that goal, ACYF recently funded a major effort to develop a more comprehensive theory and more adequate measures of the construct.

During the first two years of the project, the evaluator worked closely with local projects to identify site-specific program objectives and possible measures of these objectives. At the end of Year II, the decision was made, in consultation with ACYF, not to invest further resources in the development of a site-specific battery tailored to local program emphases. Among the factors taken into consideration in making that decision were (1) the additional cost of collecting site-specific data, (2) the difficulty of making operational distinctions among the objectives of local programs, and (3) the difficulty of finding or developing suitable measures. Readers familiar with the national evaluation of Follow Through (e.g., U.S.O.E. [W. Haney], 1977, pp. 179-202) will remember that a similar decision was made in that study, for similar reasons. Efficient, reliable, and valid measurement of outcomes other than specific academic achievement as defined by standardized achievement tests has long been and remains an elusive goal.

Discussion of particular measurement procedures will be taken up in Chapters III through VI as appropriate. They include a broad range of both qualitative and quantitative measures of community factors, organizational and economic features of centers and schools, demographic characteristics of major participants, classroom social structure and instructional process, and attitudes, knowledge, and behaviors of teachers, parents, and children.

Certain peculiarities in the data collection design are worth noting and explaining here, before the reader encounters them in the text that follows. The baseline data on parents and teachers will seem curiously skimpy. From 1974 through 1978, ACYF and the evaluator engaged in a running battle with the U.S. Office of Management and Budget (OMB), which had been given new responsibilities for reviewing instruments to be used in human research and a specific presidential mandate to reduce the burden that federally funded research placed upon its subjects. Readers who were also federally supported researchers during this period will remember how difficult it was to obtain OMB clearance for newly developed measures that would be administered to "more than nine" subjects.[1] For example, because OMB clearance could not be obtained for standardized interview schedules, the Year III (1976-77) implementation study (Smith et al., 1977) could only be conducted at "nine" of thirteen sites by interviewing a total of "nine" principals, "nine" teachers, "nine" parents, and so forth across all sites. OMB's failure to provide timely clearance for proposed measures (in some instances over a period of years) very nearly brought the evaluation to a complete halt. As a consequence of OMB constraints on data collection, little information is available on parents or teachers prior to 1979, by which time children in the evaluation cohort were already in first grade.

ATTRITION

Before discussing the approach to data analysis taken in this volume, it is necessary to describe the ravages of attrition, which have required serious and repeated rethinking of the original longitudinal analysis plan.

Attrition is surely the worst enemy of longitudinal research. It affects not only the subjects of research but the researchers themselves. The acknowledgments at the beginning of this volume show clearly the dramatic effects of attrition on the evaluator's staff from 1974 through 1982. Fortunately, we experienced attrition with replacement so that the

[1]The clearance system and associated problems are nicely illustrated by case studies in an article prepared by Launor Carter (1977).

evaluation persisted. However, major credit for what "evaluational continuity" has been achieved must go to our ACYF Project Officer, one of the few persons to have been actively involved in PDC from its conception to the present, having survived even the recent disbanding of ACYF's Research and Evaluation Branch and massive reductions in the federal work force.

The PDC sample has not fared as well. Losses have occurred for all program and analysis units--sites, institutions, staff, parents, and children. The attrition has been of three major types:

- Program attrition. Program attrition occurred when a site, a school within a site, a teacher, or a family ceased to participate in PDC. Sites dropped out of the program when contracts with ACYF were not renewed. PDC schools within a site dropped out when school administration opted out of program participation. Comparison schools dropped out of the program when no Comparison group members of the evaluation cohort were enrolled; through 1979, Comparison schools dropped in when a Comparison child did enroll.[1] Teachers dropped out or in depending upon the school in which they were teaching. Parents and children dropped out of PDC when children moved to non-PDC schools. Parents and children dropped out of the Comparison "program" when children moved to a PDC school or to a non-PDC school not included in the "matched" and "supplemental" pool identified in 1979.

- Sample attrition. Attrition from the program always resulted in attrition from the evaluation sample. In addition, sample attrition occurred when two sites, which remained in the PDC program, were excluded from the evaluation sample: one, because a satisfactory local Comparison group could not be found given the PDC sample's unique demographic characteristics; the other, because the local cultural-linguistic context seemed to require an entirely different approach to measurement from that used elsewhere. Sample attrition also occurred when families were "lost" either because of emigration or simple disappearance, perhaps as the result of a change in surname. Finally, when forming the final analytic sample, decisions were made to exclude 10 children, and consequently their parents, from the evaluation sample for reasons of repeated retention in grade (1 child) and first-language difference (9 children at two sites).

[1] In order to control data collection costs a decision was made in 1979 to limit the non-PDC schools to which Comparison group children would be followed to those schools within a district in which Comparison children were enrolled at that time. For purposes of institutional analysis these non-PDC schools were surveyed and classified as "matched comparison schools" (most similar to PDC schools) or "supplemental comparison schools" (all the rest).

- Measurement attrition. Measurement attrition refers to the occurrence of missing data for sites, institutions, staff, parents, or children who are counted as members of the evaluation sample. Measurement attrition resulted from such things as oversight or invalid measurement by field data collectors, absences from school among students, actual loss of protocols, and occasional refusals to participate in data collection.

When the longitudinal phase of the evaluation began in 1979, the sample of sites had been reduced from 15 to 11. Three sites had been lost through program attrition; one, through sample attrition by exclusion from the evaluation sample. The number of sites in the longitudinal evaluation was further reduced, from 11 to 10, by exclusion of a site where longitudinal data were collected but where there was no local Comparison group.

The total number of PDC and Comparison children entering Head Start as members of the longitudinal evaluation cohort at ten sites in the fall of 1976 was 1,090. Attrition from this sample is illustrated in Table II-1, by site by group by year. Overall attrition from program entry to formation of the longitudinal analytic sample[1] was 70%, or 5 to 10% higher than projected before the longitudinal phase of the evaluation began (Granville et al., 1976, p. 86). Attrition was higher in the PDC group (73%) than in the Comparison group (67%). Across sites, PDC attrition rates ranged from 52% to 85%, while Comparison rates ranged from 28% to 83%. It was anticipated that Comparison group attrition rates would be somewhat lower than PDC rates on the average since Comparison children were followed to a rather large number of non-PDC schools.

The average number of children per group per site in the longitudinal analytic sample is 16.3, but ranges from 3 children in the PDC group at Site 2 to 31 in the PDC group at Site 8. It should be emphasized that these numbers do not reflect the effects of *measurement attrition* from year to year. For most child measures and all parent measures at all testpoints the actual samples entering into analyses are somewhat smaller than the numbers reported in Table II-1 for the longitudinal analytic sample due to occasional missing data.

In a word, the original sample of children was devastated by program attrition, calling into question a fundamental premise of PDC--that it would be possible to control the environmental continuity that children experience from Head Start through the early elementary grades. Contemporary families, even those living in or at the margins of poverty, are mobile, and most children apparently cannot be counted upon to remain in continuous educational environments for very long.

---

[1]The longitudinal analytic sample comprises all children who remained in the evaluation sample through third grade. A small number of children who participated for the duration were excluded from the final analytic sample for reasons already mentioned in the preceding paragraph.

Table II-1

Summary of Year-to-Year Attrition
from Longitudinal Sample

	Group	Fall 1976 Entry	Spring 1977 Head Start	Spring 1978 Kindergarten	Spring 1979 First grade	Spring 1980 Second grade	Spring 1981 Third grade	Longitudinal Analytic Sample	% Attrition from Fall 1976 to Longitudinal Analytic Sample (81)
Site 1	PDC	44	37	25	21	17	15	13	70%
	Comp	40	32	28	25	21	18	13	68%
Site 2	PDC	55	51	25	20	15	13	8	85%
	Comp	32	27	24	14	13	9	9	72%
Site 3	PDC	56	53	32	28	22	20	18	68%
	Comp	57	56	51	38	33	31	30	47%
Site 4	PDC	47	39	19	16	14	13	13	72%
	Comp	39	38	35	31	30	28	28	28%
Site 5	PDC	50	47	25	15	11	10	10	80%
	Comp	54	51	25	15	14	9	9	83%
Site 6	PDC	44	42	32	27	24	17	17	61%
	Comp	58	49	33	22	14	10	10	83%
Site 7	PDC	66	60	31	21	18	12	11	83%
	Comp	64	62	52	33	28	15	15	77%
Site 8	PDC	64	59	49	38	33	31	31	52%
	Comp	57	55	46	41	39	30	30	47%
Site 9	PDC	48	47	29	23	15	13	13	73%
	Comp	81	76	60	39	29	21	21	74%
Site 10	PDC	58	52	40	25	18	10	10	83%
	Comp	76	68	52	30	24	17	17	78%
All Sites	PDC	532	487	307	234	187	154	144	73%
	Comp	558	514	406	288	245	188	182	67%
	Total	1090	1001	713	522	432	342	326	70%

The Consequences of Attrition

The effects of attrition on the evaluation have been quite serious. Four major consequences are discussed briefly here.

Nonequivalent Control Groups

In quasi-experimental research there is always the very real possibility that the control group will differ from the experimental group with respect to pretreatment characteristics that influence later outcomes used to evaluate treatment effects. This will be true in spite of efforts to select control subjects who resemble experimental subjects in their pretreatment characteristics and in spite of any demonstration that the groups are, in fact, similar with respect to some number of measured characteristics, for one can never be certain that all pretreatment characteristics relevant to later outcomes have been measured, or adequately measured. More disconcerting than this omnipresent threat to the internal validity of quasiexperimental designs is the frequently encountered situation in which treatment and control (or comparison) samples are demonstrably different with respect to measured pretreatment characteristics that influence outcomes. Under such circumstances, differences in outcome values are known to reflect not only differences in treatment but differences resulting from treatment-unrelated factors as well, which is to say that observed differences in outcomes are biased estimates of treatment effects. And only by ascertaining the magnitude and direction of such bias can one isolate the effects of treatment on outcomes.

In the case of PDC, the initial samples of PDC and Comparison children appeared to be quite similar both overall and within sites. However, sample attrition from Head Start through third grade gradually altered this picture. Although aggregate PDC and Comparison samples constituting the longitudinal analytic sample (n=326) remained quite similar with respect to pretreatment characteristics, many site-level PDC and Comparison samples did not. The nature and extent of pretreatment differences between site-level samples are described in detail in Appendix B--Attrition from the Longitudinal Sample. Here, it is sufficient to note that statistically significant PDC/Comparison differences were found for demographic/background characteristics at seven sites and for entry-level test scores at nine of the ten sites considered in the longitudinal evaluation. Moreover, pretreatment differences frequently appeared to affect outcome levels.

Thus, *attrition made it necessary to implement data analytic methods that would make adjustments for bias arising from control group nonequivalence.*

Low Statistical Power

As already indicated in Table II-1, site-level program samples in the longitudinal sample were quite small, ranging from 8 to 31 with a mean size of about 16. The power of statistical tests to detect differences between groups diminishes as the size of the groups diminish: the smaller the sizes of two samples, the larger must be the magnitude of program effects to achieve statistical significance at a given alpha level. For example, in order to find (Student t test) a statistically significant difference at the $\alpha = .05$ level between two samples, each containing 16 subjects (the average size of longitudinal site-level samples), the magnitude of difference in sample means must be 0.75 standard deviation units or larger. When one is interested in making reliable inferences about the population, rather than the study samples (see Cohen, 1969), even larger effect sizes are required.

Readers familiar with the findings of evaluations of large-scale educational interventions will recognize that effect sizes of 0.75 standard deviation unitsor more are rather uncommon, at least in analyses of program effects on such child outcomes as reading and math achievement. And most educators would agree that substantially smaller effect sizes can be socially or educationally significant, and very much worth detecting in a program evaluation. Although comparisons of aggregate PDC and Comparison samples (n=144 and n=182, respectively) would not be hampered by low statistical power, the existence of substantial planned and unplanned variation in PDC programs across sites necessitated site-level analyses to elucidate program effects.

Thus, massive attrition made it necessary to devise a data analytic strategy that would maximize statistical power in order to detect important but moderate effects in small site-level samples.

Limited Possibilities for Multivariate Analysis

One of the ways of increasing the power of statistical tests when faced with small samples is to employ multivariate analytic designs, multivariate with respect to independent variables (multi-faceted analysis of variance and analysis of covariance) and with respect to dependent variables (repeated-measures and MANOVA designs). Multivariate designs also make it possible to address hypotheses and to make adjustments for bias arising from group nonequivalence that are not possible in univariate designs incorporating a single independent and single dependent variable.

Attrition severely limited the possibilities for implementing analytic designs incorporating multiple dependent or independent variables. The most obvious limitation was on the available degrees of freedom, which prevented applying even modest multi-faceted analysis of covariance designs at the site level and constrained the development of more complicated designs for aggregate, cross-site analyses. A more subtle problem became apparent during interim stages of the evaluation when it was discovered that the pattern and

amount of missing data (across measures obtained at the same point in time and across time for particular measures) produced by measurement attrition dramatically reduced sample sizes in aggregate multivariate analyses of repeated measures and of different measures obtained at the same point in time. Disregarding covariates and including only major child outcome measures from the spring 1981 battery, a cross-sectional MANOVA reduced the longitudinal analytic sample by 34%--from 326 to 214 children having scores for all measures. Repeated-measures analyses of variance spanning the Head Start through third grade years, for measures available over that entire period, reduced the longitudinal sample by 29% to 47% depending upon the dependent variable considered.

The massive reductions in sample size entailed by MANOVA requirements that no subject have missing data on any dependent variable meant that, in lieu of estimating missing data, some sites would for all intents and purposes drop out of the longitudinal evaluation. Moreover, differences in the patterns of missing data from year to year and from measure to measure meant that the samples considered in different analytic designs represented sometimes very different subsamples of the total longitudinal sample. Under these circumstances, it became increasingly difficult to integrate findings from different analytic designs without either estimating missing values or reducing the analytic sample to only those subjects with all measures at all data collection points--a ridiculously small 28% subsample (n=90).

Thus, *measurement and sample attrition severely limited the possibilities for implementing analytic designs incorporating multiple independent and or dependent variables.*

Questionable External Validity

The PDC evaluation was undertaken with the hope of determining whether developmental continuity programs might be of general benefit to Head Start children. This is to say, there was a desire to generalize from the findings of this limited experiment to the Head Start population as a whole. Since project sites were not specifically selected to represent the full range of Head Start communities, generalizations beyond the communities represented would have to have been qualified under the best of circumstances, but they would not have been unreasonable.

However, the heavy attrition experienced over the course of the evaluation raises serious questions about the external validity of the research design. Given overall sample attrition of 70% (ranging from 28% to 85% for groups within sites) and further reductions of actual analytic samples due to measurement attrition, one must wonder whether site-level samples remaining in the longitudinal evaluation really represent the Head Start populations in their communities. One may even legitimately question whether or not the remaining sample adequately represents the initial evaluation sample. For example, evidence is presented in Appendix B which suggests that the children

remaining in some site-level longitudinal samples were significantly different in certain respects from children who dropped out. But statistical tests aside, can the eight children remaining in the PDC sample at Site 2 at the end of the evaluation reasonably be viewed as representing the 55 children who enrolled in PDC in the fall of 1976?

Attrition has made generalization of findings from the evaluation sample to the population(s) of interest a precarious undertaking.

EVOLUTION OF A DATA ANALYTIC STRATEGY

The ultimate challenge to the authors of this report has been to devise a satisfactory data analytic strategy for addressing the research questions stated at the beginning of this section. Considering the vast but inevitably insufficient amounts of qualitative and quantitative data that had accumulated over the evaluation's five years, the much smaller than hoped for site-level samples, the great social-cultural diversity among a small number of sites, the complicated patterns of variation among local PDC programs, the biasing effects of heavy sample attrition, and the fragile quasiexperimental design with which we had embarked upon this evaluation, we could not hope for a simple textbook solution to our problems. In fact, had we taken textbook admonitions literally we surely would have walked away from the data analysis task long ago.

Few evaluations, even the most tightly controlled, produce findings that scientists consider *conclusive;* there is always another counter-hypothesis that might be tested. Rather, the findings of particular evaluations are most appropriately viewed as weakly to strongly *indicative*. Educational evaluations, set in the real world of schools, insusceptible to the tight controls of the laboratory, seem particularly vulnerable to threats against internal and external validity, and given the recent history of such evaluations, one would be extremely naive to expect conclusive results.

Important features of our general position and approach have been described by Jacob and Patricia Cohen (cited in U.S.O.E. [W. Haney], Vol. 5, 1977, p. 254):

> The behavioral scientist relies very heavily on the fruits of the labors of theoretical statisticians. They provide guides for teasing out meaning from data, limits on inference, discipline in speculation. Unfortunately, in the textbooks addressed to behavioral scientists, statistical methods have often been presented more as harsh straightjackets or Procrustean beds than as benign reference frameworks. Typically, a method is presented with some emphasis on its formal assumptions. Readers are advised that the failure of a set of data to meet these assumptions renders the method invalid. All too often, the discussion ends at this point. Presumably, the offending data are to be thrown away.

> Now this is, of course, a perfectly ridiculous idea from the point of view of the working scientist. His task is to contrive situations that yield information about substantive scientific issues--he must and will analyze his data. ...Data analysis accepts "inadequate" data, and is thus prepared to settle for "indications" rather than "conclusions." It risks a greater frequency of errors in the interest of a greater frequency of occasions when the right answer is "suggested." It compensates for cutting some statistical corners by using scientific as well as mathematical judgment, and by relying upon self-consistency and repetition of results. Data analysis operates like a detective searching for clues rather than like a bookkeeper seeking to prove out a balance.

As "detectives," however, we chose not to limit ourselves to numerical clues but to consider all available information believing that:

> ...man is, in his ordinary way, a very competent knower, and qualitative commonsense knowing is not replaced by quantitative knowing. Rather, quantitative knowing has to trust and build on the qualitative, including ordinary perception. We methodologists must achieve an applied epistemology which integrates both. (Campbell, 1975, p. 191)

In short, we view the PDC evaluation as indicative rather than confirmatory, and ourselves as detectives seeking to understand the processes and products of the PDC intervention as best we can by piecing together available qualitative and quantitative clues.

The original plans for quantitative data analysis proposed heavy reliance upon multivariate and repeated measures techniques, utilizing analysis of covariance to eliminate or reduce the pernicious effects of PDC/Comparison noncomparability and to increase the power and precision of tests for program effects. Numerous strategies employing multivariate analysis of covariance (MANCOVA) procedures for repeated measures and cross-sectional analyses have been experimented with in interim reports from 1976-77 (Head Start) through 1979-80 (second grade). These methods were applied exclusively to child outcome data; other quantitative data on parents, teachers, and classrooms were either not analyzed or were analyzed utilizing simple univariate analyses without covariance adjustments. All of these analyses aggregated PDC and Comparison group data across sites. Analyses of child outcomes, but not of other data, incorporated group-by-site interaction terms to identify site-level program impacts that departed from overall findings. Interim analyses were performed on the samples available at each data collection point, and the samples of children and parents in the evaluation sample grew steadily smaller over time.

The whole-sample multivariate analysis of covariance approach, combined with post hoc examination of site-level effects, was technically and practically appealing. Repeated-measures MANCOVA avoided assumptions of homogeneous covariances among repeated measures associated with conventional analysis of variance (ANOVA) procedures for repeated measures analysis and held out the hope of identifying small, but longitudinally cumulative, program effects in small site-level samples. Cross-sectional MANCOVA brought similar statistical power to tests of program effects at particular points in time through simultaneous comparison of multiple outcome variables. Because multivariate approaches promised greater statistical power than univariate analysis of variance (ANOVA) to ferret out program effects of magnitudes commonly associated with effective educational interventions, the multivariate data analytic strategy could not be dismissed lightly.

However, as the longitudinal evaluation progressed, measurement attrition (missing data) posed increasingly serious problems for multivariate analysis as described in the preceding discussion of attrition's consequences. By the spring of 1981 (third grade), the MANCOVA requirement of no missing data could only be met by enormous sacrifices in sample size. And already by spring 1979 (first grade), the patterns of findings from univariate and multivariate analyses were beginning to diverge significantly due not to differences in precision but to differences in the samples analyzed--differences which were caused by the non-missing data requirements of MANCOVA techniques. As attrition continued to take its toll, it became increasingly probable that the patterns of effects identified by multivariate analyses could not be generalized to the total evaluation sample. And as time wore on, much of the promised increase in statistical power of multivariate tests was lost because of associated reductions in sample sizes. One possibility for salvaging the multivariate approach to data analysis was considered. It involved eliminating missing data by estimating missing scores from other background and outcome data. This option was dismissed after concluding that the extent of missing data was such that we would find ourselves in the position of analyzing substantially "fictitious" data sets in many cases.

Since a full-blown multivariate approach to data analysis seemed impractical, we turned to less powerful univariate designs that would have to rely upon convergence of findings from different, and not always independent, analyses to elucidate program effects. The most significant limitation this decision placed upon the evaluation was a reduction in statistical power, power to identify small to moderate effects in very small site-level samples. In other words, it became more likely that we would make Type II errors--i.e., overlook real program-related differences in outcomes--and correspondingly less likely that we would make Type I errors--i.e., find differences that were simply due to chance.

Though we could not escape this limitation, *we hoped to mitigate the problem of low statistical power by performing somewhat less conservative tests than is customary.* At the same time, we hoped to counterbalance the

the potential increase in Type I errors resulting from the relaxation of alpha levels by insisting upon "triangulation" or internal consistency among findings from different data sources, different analysis strategies, and different impact domains defined by the intervention model. For example, we decided to interpret statistical findings in child outcomes as program-related effects only when (1) the results from multiple analytic perspectives were consistent with one another, (2) major alternative hypotheses could be discarded, (3) potentially explanatory effects were suggested by parent, teacher, and or institutional outcomes, and (4) there was no contraindication in available qualitative data or in commonsense (informed by other research, theory, and what Campbell called "ordinary perception").

Next, we considered whether or not to employ analysis of covariance when possible. Univariate analysis of covariance (ANCOVA) offers two potential advantages over simple analysis of variance (ANOVA). First, like its multivariate analogue, ANCOVA can increase the precision, efficiency, or power of tests of program effects, making it possible to identify reliable effects that are smaller than those detectable by ANOVA. When ANCOVA works properly, it removes variance in the outcome (dependent) variable that is associated with (explained by) variance in the covariates (usually background characteristics uncorrelated with treatment). This reduces the amount of variance in the outcome variable left to be explained by program differences, thus reducing the error variance (variance unexplained by program, as well as any other design facets and their interactions) and increasing the probability of finding a statistically significant program effect if such an effect is actually present--i.e., increasing the power of the test. (This scenario assumes that the variance reduction achieved through covariance outweighs any reduction in sample size due to missing data on the covariates and the loss of degrees of freedom associated with the incorporation of covariates in the design.)

Second, under the right circumstances ANCOVA can be used to adjust outcome variables so as to eliminate program or treatment group differences in outcome levels due not to treatment but to differences in background characteristics. Growing evidence of control group nonequivalence due to nonrandom attrition at some sites made this application of ANCOVA techniques increasingly attractive. Traditionally, the application of ANCOVA techniques for this purpose was considered justifiable only when treatment group samples were demonstrably drawn from the same population by random selection/assignment, in which case any group differences with respect to pretreatment characteristics would have resulted from simple sampling error. In recent years, however, ANCOVA has been used in quasiexperimental research, where it is not possible to assume that program groups represent samples from the same population, to adjust outcomes for pretreatment differences that were not due to chance but to systematic biases in initial sample formation and/or sample attrition. Quasiexperimental evaluations of both Head Start and Follow Through, as well as many other educational interventions, have used covariance procedures of one sort or another in an effort to "adjust away" differences in outcomes due to differences in pretreatment characteristics.

Although ANCOVA applications of the second sort have become accepted practice among evaluators faced with the problem of estimating program effects by comparison of nonequivalent groups, very serious questions have been raised about the effectiveness and trustworthiness of covariance adjustments under quasiexperimental conditions.[1] Lee Cronbach and colleagues put it this way in the title of an occasional paper released in 1976: "Analysis of Covariance: Angel of Salvation, or Temptress and Deluder?". Their conclusion was that analysis of covariance approaches inevitably tempt and frequently delude the practical evaluator, but that such approaches in one form or another currently offer the only possible salvation for many real world evaluations:

> The solution is not to abandon realistic social science but to make less presumptuous claims regarding the result. (Cronbach et al., 1976, p. 34)

The most disconcerting thing about using analysis of covariance to estimate treatment effects is that one cannot be certain in actual evaluations (in contrast to statistical simulations) whether or to what extent covariance adjustments have had the intended effect. When treatment effects are confounded with the effects of subject characteristics that existed prior to treatment, analysis of covariance probably never produces, and certainly could never be demonstrated to have produced, fully unbiased adjustments in outcome variables.

Though the necessary conditions for achieving unbiased adjustments vary somewhat according to the particular covariance approach employed, in general the following conditions would have to be satisfied:

- Reliable covariate. The covariates are perfectly reliable.
- Complete covariate. The covariates fully specify differences in subject characteristics that make a difference in outcome.
- Homogeneous regressions. The relationships between the covariates and the outcome variable are identical for all groups. (In Cronbach et al.'s [1976] formulation, this condition would not have to be met if the "complete covariate" or "complete discriminant" were available for each group.)

[1] See for example: Campbell and Erlebacher, 1975; Bryk and Weisberg, 1977; Lord, 1969; Cronbach, Rogosa, Floden, and Price, 1976.

All measures of important human characteristics are in some degree fallible, and covariance models frequently incorporate baseline demographic data for which not even primitive reliability/generalizability estimates are available. The complete covariate (or complete discriminant) still eludes the researcher's grasp. And fully homogeneous regressions (in contrast to regression slopes that are not significantly different) are quite improbable. Though the most common bias of covariance adjustments appears to be underadjustment, overadjustment and even adjustment in the wrong direction are entirely possible. To the extent that covariance analysis falls short of satisfying these assumptions, estimated treatment effects must be interpreted cautiously and without presumption. But, having no alternative, *we decided to utilize several analysis of covariance techniques* involving somewhat different assumptions, in conjunction with other data analytic strategies to help us glean what information we could from the PDC evaluation's quantitative data.

Two further issues had to be resolved in devising an analytic strategy: Should we analyze data aggregated across sites or unaggregated by sites? Should we treat program-by-site groups, schools, classrooms, or individuals as the units of analysis when addressing particular questions? The first question involves decisions about *level* of analysis; the second, about *unit* of analysis.

Given the PDC intervention model, which encouraged site-level program variation, and given our knowledge that substantial site-level variation had indeed occurred even in program areas that were supposed to be common across all sites, estimation of program effects at the site level, in some fashion, would be essential if any sense were to be made of measured outcomes for institutions, teachers/classrooms, parents/homes, and children. Thus, *we decided to conduct primary analyses at the site level.*

The unit of analysis question took different forms depending upon the research question being addressed and specific data being analyzed. When addressing questions of institutional impact, for example, it was no simple matter deciding what the unit of analysis was or should be. Schools and centers were natural units to examine for certain purposes, by considering data collected at the school level and by aggregating data on teachers, parents, or classrooms to the school level. However, within the PDC sample, but not the Comparison sample, we were also interested in a higher level of organization than the individual school, for the PDC intervention was intended to forge formal, structural links among participating schools and centers, thereby creating a larger institution. Thus, for some purposes the unit of analysis within the PDC sample would be the local "PDC project," a unit without parallel in the Comparison sample. When evaluating impacts on parents and teachers, the appropriate units of analysis appeared to be classrooms and

homes as well as individual teachers and parents. Though conceptually it made some sense to aggregate classrooms to the school or program level within sites, the number of units resulting from such aggregation would be too small to permit quantitative tests of site-level program effects. Regarding the unit of analysis in evaluating child outcomes, serious attention was given to the possibility of aggregating individual child data to some higher level, particularly the classroom level. This was not done for four reasons. First, the site-level samples of classrooms were quite small. Second, the number of children tested per classroom was extremely small in many cases. Third, nonrandom distribution of children across classrooms and radically unequal class cell sizes threatened to produce class means that would misrepresent actual program outcomes. Fourth, the fact that children not only changed classrooms but changed membership in class samples from year to year made it impossible to think of class samples as distinct treatment groups from a longitudinal perspective. It is worth noting that from 1978-79 through 1980-81, 44-56% of classes in the evaluation sample contained only 1 cohort child, while 71-76% contained no more than 2 children. Under these circumstances, class means were unlikely to have much meaning. Aggregation of child data to even higher levels--i.e., school or site-level program--would have presented even greater impediments to meaningful site-level analysis. In sum, *the following major units of analysis were defined: schools and centers; schools and centers aggregated within program at the site level; individual teachers and parents; classrooms and homes; and individual children.*

The preceding discussion has touched upon the analytic strategy devised for qualitative data as well as strategies for integrating qualitative and quantitative data; however, a few additional comments would seem to be in order. The bulk of qualitative data describe institutional features or institutional contexts and were collected either formally, through semi-structured interviews, or informally, through conversation and observation during the course of site visits from 1975 through 1981. These data serve primarily to document implementation of PDC Guidelines, both intentionally by PDC schools/centers and unwittingly by Comparison schools/centers. Secondarily, they document factors within local schools/centers and the larger environment that facilitated or hindered Guideline implementation. These data were in the form of descriptive statements, written on paper and stored in file cabinets.

The qualitative data analytic strategy was developed to address research questions stated at the beginning of this section and to mitigate threats to internal and external validity of the same sort that affected the quantitative analysis--nonequivalent samples, nonrepresentative samples, measurement unreliability, incomplete specification of relevant characteristics of the particular unit of analysis, and so forth. Specific methods for qualitative data analysis are described in Chapter III--PDC's Influence on Local Institutions--of this volume and in Volume II.

The final analytic step in the evaluation required inferences about the effectiveness of the intervention at all levels of the theoretical model--institutions, teachers/classrooms, parents, and children, utilizing all available qualitative and quantitative information. This step in any evaluation is necessarily qualitative even when based exclusively on the findings of quantitative analyses, and whether the evaluator is dealing with a single finding from one grand multivariate analysis or a muddled pattern of findings from many independent or interdependent tests. In the end, the evaluator and the reader must rely upon logical analysis and qualitative inference, informed and tempered by theory, the findings of other research, and commonsense.

DATA ANALYSIS METHODS USED IN THE FINAL EVALUATION

The specific data analysis methods used in this final stage of the evaluation were selected and devised to implement the general strategy described in the preceding section. They are the result of much exploratory work and compromise. Although numerous other approaches might legitimately have been taken (and various alternatives were, in fact, tried and evaluated before final decisions were made), we are fairly confident that other approaches would not produce substantially more valid conclusions.

The methods described in this section were used to analyze all parent, teacher/classroom, and child outcome data. Measures, findings, and any particular variants in the data analytic designs applied to these three outcome domains are reported in Chapters IV (Parent Outcomes), V (Teacher/Classroom Outcomes), and VI (Child Outcomes). Methods used to analyze qualitative and quantitative data describing implementation of institutional features prescribed in the PDC Guidelines are discussed in Chapter III--PDC's Influence on Local Institutions. Analyses of Guideline implementation did not involve statistical tests, but only the production, aggregation, and "qualitative" comparison of ordinal scale ratings.

Design Criteria

Development of analytic methods for the longitudinal evaluation was guided by the following considerations:

- Primary analyses would be conducted at the site level; secondary analyses comparing aggregate PDC and Comparison samples would be interpreted in light of site-level findings. (This approach departs from designs used earlier in the evaluation which performed post hoc comparisons of site-level samples following aggregate tests indicating overall effects.)

- Only univariate designs would be employed--i.e., each outcome measure at each data collection point would be analyzed separately.

- Alpha (significance) levels would be relaxed in order to compensate for low statistical power in site-level analyses; however, steps would be taken to control Type I error rates by insisting upon "triangulation" of findings from different analytic designs, for related outcomes, and across levels of the design defined by the intervention model.

- Analysis of covariance procedures of some sort would be employed both to increase the power of univariate tests and to attenuate, if not eliminate, bias in tests of program effects resulting from PDC/Comparison non-equivalence at the site level.

- Given the uncertain validity of analysis of covariance assumptions under conditions of group nonequivalence, at least two designs (based on rather different assumptions) would be employed, and convergence of findings from the two designs would be taken as evidence of probable program effects.

Our reasons for establishing these criteria have already been presented.

After lengthy exploration and review, two analytic methods were selected for evaluating PDC/Comparison differences on all parent, teacher, classroom, and child outcomes. In the remainder of this report, these methods are referred to as Designs 1 and 2. In addition, procedures were developed for synthesizing the findings from these two designs and for identifying patterns of findings across time and across related outcome variables. Descriptions and discussions of all methods follow.

Design 1

This is certainly the less orthodox of the two approaches. Although its component elements are familiar, the total package has no clear precedent in the evaluation literature with which we are familiar.[1]

[1]While bearing some relation to the so-called "Belson method" (Bryk & Weisberg, 1977; Cochran, 1968), Design 1 departs from that method in significant respects.

Design 1 rests on the assumption that outcomes in the aggregate Comparison sample provide the best available estimates of what the outcomes would have been in the aggregate PDC sample in the absence of the PDC treatment. Such an assumption would be unjustifiable if aggregate PDC and Comparison samples were very different with respect to pretreatment characteristics; however, available evidence suggests that the aggregate samples, if not the site-level samples, were quite similar (see Appendix B) and that even in the presence of the PDC intervention, pretreatment characteristics correlated similarly with outcomes in the two aggregate samples (see results of aggregate tests of homogeneity of regressions in Appendix I). Design 1 does not assume, nor does available evidence (Appendix B) support an assumption, that PDC and Comparison site-level samples would have evidenced the same outcomes in the absence of the PDC intervention.

The basic design involves comparing the results of three tests to determine whether or not a program effect is likely. Each of these tests is described below:

1. Comparison of Predicted Scores. Regression equations derived from aggregate Comparison group data are used to predict PDC outcomes as they would be expected to have occurred in the absence of the PDC intervention. The multiple linear regressions performed on aggregate Comparison group data incorporate as covariates major measures of subjects' pretreatment characteristics that have been demonstrated to predict outcome variables. The covariate sets used in analyses of parent, teacher/classroom, and child outcomes were rather different; they are described in Chapters IV, V, and VI, respectively. Predicted outcome scores for PDC and Comparison samples were compared in Student t tests. Information about subjects' site membership was not included in the regressions in spite of evidence that outcome levels and covariate-outcome relations varied by site. Including "site" as a blocking factor in the regressions contributed very little to predictions over and above the contributions of the covariates, which is to say that the covariates largely captured differences among site samples that influence outcome levels. There was one notable exception to this generalization: aggregate regressions fairly consistently generated predicted scores that were much higher than observed scores for both PDC and Comparison children in one site where children were predominantly Hispanic and Spanish-dominant at program entry. Predictions for Comparison children in that site were dramatically improved by incorporating information about initial language dominance in the regression equation; given the fact that no other Spanish-dominant children remained in the longitudinal sample, this solution amounted to introducing a single site dummy variable contrasting children at the site in question with all other children in the sample.

Evidence of differences in covariate-outcome relations across sites would suggest including interactions of "site" with covariates in the regression design. This was not done for two reasons. First, there were simply not enough degrees of freedom available to interact nine site dummy variables with a fairly large number of covariates. Second, even if it were possible, we did not believe that it would be advisable given the very small site-level samples that would be used to generate regression coefficients. It seems highly unlikely that the covariate-outcome correlations found in site-level program samples, PDC or Comparison, provide estimates of population values that are as reliable and valid as those derived from analyses of the aggregate Comparison sample disregarding site. Supplemental analyses revealed patterns of covariate-outcome relation at the site level that were frequently very unstable over time and unbelievably different for the two program groups at particular points in time. Frequently, temporal shifts in relation or dramatic between-group differences were the product of one subject's scores. Unfortunately, with very small samples, not much can be done to resolve the problem of outliers.

2. Comparison of Observed Scores. The outcome scores actually obtained by the PDC and Comparison groups were compared in a Student t test.

3. Comparison of Predicted with Observed Scores in Each Group. Predicted and observed scores were examined for each group to determine whether predicted scores fell within, above, or below a confidence interval placed around observed scores.

Given the low power of statistical tests conducted at the site level, we relaxed the alpha (significance) level required for accepting a difference as indicative of a program effect to .10 for comparisons of observed scores and comparisons of predicted with observed scores (.90 confidence interval). However, we decided to set alpha at .05 for comparisons of predicted scores. Our reasoning was that we should compensate for the much greater power that tests of predicted scores had relative to tests of observed scores. This greater power was due to the fact that predicted scores had much smaller variances than observed scores. For the Comparison group, predicted scores contained only that variance accounted for by the covariates, which never explained more than 50% of the variance in observed scores, and typically explained much less. Selection of .05 rather than some other level for tests of predicted scores was arbitrary but "feels" right having lived for some time with the consequences. This decision did not, insofar as we could determine, favor either group over the other. In analyses comparing the aggregate PDC and Comparison samples, alpha for comparisons of observed scores and of observed with predicted scores was set at .05; alpha for comparisons of predicted scores, at .01.

Even when alpha is relaxed to the .10 level, however, the power of Student t tests of observed site-level means is quite low. For example, in order to find a statistically significant difference between PDC and Comparison samples each containing 16 subjects (the average size of longitudinal site-level program samples) the magnitude of difference between sample means must be .62 standard deviation units or more. An educational evaluation requiring differences of this size (particularly for child outcomes like academic achievement) to demonstrate program effects is unlikely to reach strong conclusions about the effectiveness of the program being evaluated. Unfortunately, relaxing alpha levels even further would not improve matters since the frequency of Type I errors would increase rapidly, casting serious doubt on all reported findings of program effect.

Synthesizing the Results of Four Tests

Synthesizing these test results into a summary effect statement is accomplished by logical analysis, applying a set of decision rules. The complete set is presented in Appendix I together with tables of the first-order results to which they were applied. First, findings from tests of predicted scores are integrated with findings from tests of observed scores. The first-level inferences associated with all combinations of effects are illustrated in Table II-2. One combination leads to a conclusion that program effects are improbable--when no difference was predicted and none was found. Four combinations lead immediately to the inference that program effects are probable--when no difference is predicted, but one is found, and when a difference is found in the opposite direction of that predicted. Four other combinations raise the possibility that there are program effects--when predicted differences are not found and (less likely) when predicted differences are found.

Second, further distinctions are made among "possible" effects by taking into account the findings of predicted observed score comparisons. These findings indicate whether one or the other or both groups performed differently than predicted. (In the case of aggregate tests, the mean of Comparison group predicted scores is, of course, always identical to the mean of their observed scores given the regression design.) Table II-3 gives some examples of how this information is used to differentiate possible effects from step one into "improbable" (0), "possible" (+? or -?), and "probable" (+ or -) summary statements of effect. The end result of this procedure is to assign to each group comparison on each measure at each site one of the following values: +, +?, 0, -?, or -.

Table II-2

Illustration of Decision Rules for Synthesizing Findings from Design 1 Analyses: Step One

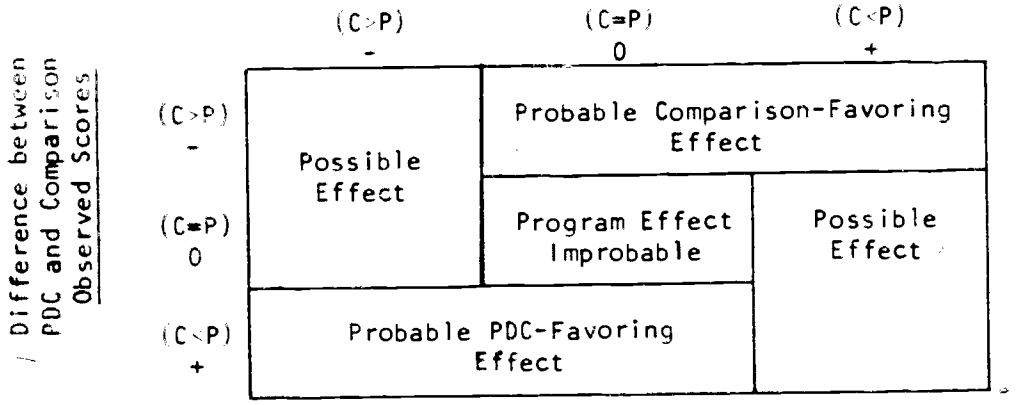


Table II-3

Illustration of Decision Rules for Synthesizing Findings from Design 1 Analyses: Step Two[1]

PDC-COMP Predicted	PDC-COMP Observed	PDC: Observed-Predicted	COMP: Observed-Predicted	Effects Summary	Interpretation
+	+	0	0	0	PDC-favoring effect predicted and found. Program effect improbable.
+	0	0	0	-?	PDC-favoring effect predicted but not found. Comparison-favoring effect possible.
+	0	-	0	-	PDC-favoring effect predicted but not found because PDC scored significantly lower than predicted. Comparison-favoring effect probable.
+	+	+	0	+?	PDC-favoring effect predicted and found; however, PDC appears to have scored even higher than expected while Comparison did not. PDC-favoring effect possible.

[1]The complete decision rule set is presented in Appendix I.

Design 2

Design 2 will be familiar to readers of the educational research literature--two-way analysis of covariance (ANCOVA) incorporating "group" (P/C) and "site" as design facets, and predictor variables from Design 1 as covariates. For purposes of this evaluation, group effects within sites were tested first by including 10 group-by-site interaction terms in a multiple linear regression, together with nine site dummy variables and all covariates, but not "group." A secondary analysis of the aggregate group main effect was performed in a separate regression incorporating a group dummy variable, nine site dummy variables, and all covariates, but no group-by-site interaction terms. For site-level tests alpha was set at .10; for aggregate tests, at .05. Differences significant at these levels were judged to indicate probable (+ or -) program effects; "possible" effects were not identified in this approach.

In Design 2, covariance procedures are used to adjust outcome scores before group comparisons are made. It is not assumed that covariate-outcome relations in the Comparison group best represent the relations that would hold in the PDC group in the absence of treatment. Instead, it is assumed that the best estimates of covariate-outcome relations (regression coefficients) are pooled estimates representing weighted averages of within-group regression slopes. If within-group covariate-outcome correlations are quite different (i.e., if regressions are heterogeneous), the adjustments in outcome scores made by covariance analysis may be quite misleading--failing to eliminate bias or even increasing bias due to group nonequivalence. If within-group regressions are very similar and if covariates strongly predict outcomes, analysis of covariance may (see the preceding section of this chapter for a more detailed discussion of ANCOVA issues) go a substantial way toward removing variance in outcome measures associated with differences in pretreatment characteristics (i.e., bias) and increasing the power of statistical tests (by reducing "error" variance in tests of between group differences).

Because the ANCOVA design incorporated more covariates than there were subjects in some group-by-site samples, it was impossible to compute customary tests of the homogeneity of within-group (i.e., within-group, within-site) regressions. And supplemental analyses, already mentioned in the preceding discussion of Design 1, raised serious doubts about assuming (in the absence of a direct test) that regressions were homogeneous, or sufficiently homogeneous not to pose a significant threat: bivariate correlations and simple multiple regression models relating covariates with outcomes within-groups within sites frequently revealed very different covariate-outcome relations between site-level groups. However, these hints that full-scale regressions would be dissimilar between site-level groups should be tempered by the other observation made in that section--specifically, that correlations and regression coefficients derived from our small site-level samples cannot be trusted to represent corresponding population values. The pooled estimates of regression coefficients used in Design 2

are very likely more valid than would be estimates based on within-group, within-site regressions, assuming the latter regressions were possible, as they would be if much simpler ANCOVA designs were used. Whether the ANCOVA adjustments accomplished by Design 2 were adequate and appropriate (i.e., permitted unbiased tests of PDC/Comparison difference) is another question, one that cannot be answered with available information.

Synthesis of Findings from Designs 1 and 2

Each design has potential weaknesses, the implications of which cannot be fully known. By applying both designs and looking for convergence of their findings, we hoped to achieve higher levels of confidence in our conclusions about program effects than would be possible were we to rely upon either approach by itself. The following summary effects were generated by the synthesizing process:

+	probable PDC-favoring effect
+?	possible PDC-favoring effect
+??	less possible PDC-favoring effect
0	no program effect
-??	less possible COMP-favoring effect
-?	possible COMP-favoring effect
-	probable COMP-favoring effect

The decision matrix for synthesizing findings from Designs 1 and 2 is presented in Table II-4. The shaded areas indicate contradictory findings (+/-) from the two designs. If such findings appeared with any frequency, the entire approach would be called into question. Fortunately, in practice, contradictory findings were extremely infrequent (and coded "0" when they did occur).

Findings for all Design 1 and 2 tests and summary effects are presented in Appendix I by site and overall. Also reported there are group-by-site and aggregate group observed, predicted, and ANCOVA-adjusted means for all outcome variables.

Identification of Patterns of Effects

After much effort and deliberation the total number of measures considered in the longitudinal evaluation of parent, teacher/classroom, and child outcomes was whittled down to 49 individual and composite indicators--no more than 10% of the total number of discrete measures available.

Table II-4

Decision Matrix for Synthesizing Findings
from Designs 1 and 2

		Design 2 Effects		
		+	0	-
Design 1 Effects	+	+	+?	
	+?	+	+??	
	0	+?	0	-?
	-?		-??	-
	-		-?	-

Since many measures were repeated over time and since the analytic approach was univariate, the total number of variables analyzed in evaluating parent, teacher/classroom, and child outcomes was 96. And since results are reported for each site and for the aggregate PDC and Comparison samples, the total number of summary effects with which we have to contend is over one thousand, while the number of statistical tests exceeds five thousand.

The reader will be relieved to know that we do not intend to discuss each of these findings, or even each of the possible/probable effects. Clearly in an evaluation of this scope the occasional program effect is inconsequential. What merits our attention are patterns of program effects at particular sites, patterns that emerge over time and across related outcomes. The great advantage of multivariate analytic designs is that they are intended to detect just such patterns in either repeated measures or multiple dependent variables measured cross-sectionally. Thus, our decision to abandon a multivariate approach to data analysis was not taken lightly. But having abandoned this approach, we were compelled to develop other means for pattern identification.

As a first step toward pattern identification, we grouped outcome measures at different levels of the evaluation design--parent/teacher/classroom/child--into *domains* comprising different measures that seemed to tap important dimensions of the same construct--e.g., "parent involvement," "educational management," and "specific academic achievement." The classification process was guided by the PDC concept and objectives, social and psychological theory, research findings, commonsense and, finally, empirical evidence of relationship. We did not insist that all measures assigned to the same domain exhibit strong, positive intercorrelations with one another. To have done so would have meant, for example, treating measures of children's social adjustment in test situations separately from measures of adjustment in the classroom situation, since these two measures were not highly correlated. But theory and commonsense suggested that these variables measured different aspects of the same underlying construct, and probably *should* have correlated more strongly than they did. Most probably the very weak correlation found between these two variables was a product of their relative unreliability. Both measures were ratings--made, in one case, by testers and, in the other case, by teachers. As measuring instruments, teachers in particular seem to have been calibrated rather differently from one another. But even if these ratings had been perfectly reliable and valid, we would not have necessarily expected strong, much less perfect, correlation. Social interactional competence is in some measure context-bound; persons are not equally competent or incompetent in all social situations. Therefore, insofar as these and other measures seemed to tap some important facet of the same construct and insofar as they did not correlate inversely with one another in the total sample, we assigned them to the same domain.

Having accomplished this grouping, we next sought to develop a method for identifying outcome domains in which there were "consequential" patterns of effects at particular sites. After wrestling for some time with the

temporal dimension of effects for measures that were repeated over the course of the evaluation, we concluded that this dimension was not generally important. That is to say, the relative frequency of possible/probable effects did not seem to vary systematically over time. Thus, we decided not to treat outcome domains as two-dimensional (measure-by-time) matrices, but as effect aggregates or non-dimensional effect pools with potential directionality--favoring either PDC or the Comparison group.

Deciding whether the pattern of effects (zero/plus/minus) in a particular domain evidenced sufficient strength and directionality to be considered "consequential" demanded judgments that were more "aesthetic" and "commonsensical" than statistical. The most important defining characteristics of "consequentiality" seemed to be *internal consistency* and *strength*, and any set of decision rules would have to take both of these features into account. Of particular concern was whether the addition of *possible* (?) and *less possible* (??) effects to *probable* effects altered directionality or directional strength. Since the measures included in a particular domain were to some extent interdependent, one would anticipate high internal consistency of effects. And assuming that the intervention actually had appreciable and consistant impacts, one would expect the number of effects not only to increase as less probable findings were admitted, but to increase in directional strength. If the addition of possible (?) and less possible (??) effects did not increase the directional strength of effects, we would be forced to conclude that our efforts to compensate for low statistical power had simply introduced noise. And if their addition reversed the directionality evident in probable effects, we would have to acknowledge the possibility that our relaxation of alpha (significance) levels had introduced bias.

The decision rules finally selected were the simplest of those we tried and evaluated. They appeared to measure "consequentiality" as well as any and had two major advantages. First, they could be communicated easily to readers of the evaluation report. Second, very similar rules had been applied in another national evaluation--the evaluation of Follow Through conducted by Abt Associates (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977, p. 133). In addition, they avoided giving a false impression of precision at a point in the analysis where statistical precision was simply not possible.

Method for Identifying Consequential Patterns of Effect

The strength and direction of effects within each domain were represented by *net + or - effects as a proportion of the number of tests conducted within the domain.* Positive (P>C) effects were assigned a value of +1; negative (C>P) effects, a value of -1; and null (P=C) effects, a value of 0. The mean of these values was computed for each domain, then divided by the number of values, which equals the number of tests conducted, in the domain. Net effects computed in this manner ranged in value from -1.0 (all negative findings in domain) to +1.0 (all positive, PDC-favoring findings in domain).

Given the low power of statistical tests conducted in site-level analyses, we were anxious to consider all findings from the synthesis of Designs 1 and 2--*possible* (?) and *less possible* (??), as well as *probable*, effects--when assessing the "consequentiality" of patterns of effects within domains. The most conservative approach would restrict estimates of net effects to analyses of only *probable* findings (those confirmed by both Designs 1 and 2); the most liberal approach would derive estimates from analyses of all findings--*probable*, *possible*, and *less possible*. Hoping to steer a safe course between these extremes, which threatened excessive Type I error on the one hand and excessive Type II error on the other, we elected to derive primary estimates of net effects from analyses of pooled *probable* and *possible* effects. These effects represented PDC/Comparison differences found to be statistically significant at the .10 level or better in Design 1 and/or Design 2 analyses conducted at the site-level (.05 or better in aggregate tests). Only when "consequential" net effects were found in this analysis would we go on to consider the contribution of *less possible* findings to the net effect in a particular domain. Furthermore, we decided to include *less possible* effects in final estimates of net effects only when they increased the directional strength of effects found in primary analyses of *probable* and *possible* effects.

In addition to estimating the net observed effect for each domain at each site, the net predicted effect was also determined. This involved applying the procedure used to compute net observed effects to the findings of Design 1 tests comparing PDC with COMP scores predicted from multiple linear regressions of outcomes on pretreatment characteristics within the aggregate Comparison sample. Positive findings from these tests indicated a PDC-favoring bias in pretreatment characteristics; negative effects, a Comparison-favoring bias. Although both Designs 1 and 2 were intended to control for bias resulting from initial nonequivalence of the samples, we could not be certain that bias had in fact been eliminated. Thus, the net observed effect for each domain was interpreted in light of the net predicted effect as a final precaution against erroneous inference of possitive or negative program impacts from comparisons of biased outcomes.

The last step in identifying consequential effects involved setting a cutoff to differentiate "consequential" from "inconsequential" findings--i.e., to differentiate findings that we would discuss and attempt to explain from those that we would not. Since the net effect scale ranged from -1.0 to +1.0, both negative and positive cutoffs were required, falling somewhere between the zero midpoint and either end of the scale. But being unable to determine (at this level of synthesis) the probability with which particular magnitudes of net effect might be expected to have occurred by chance, and being unable to cite convention, we ultimately made an arbitrary decision that "looked" and "felt" right, setting cutoffs at -.25 and +.25. When

net positive or negative effects from analyses of *probable* and *possible* (?) findings equaled or exceeded 25% of the potential number of effects in a domain, we considered the pattern of effects for that domain to be of sufficient consequence to warrant specific attention in our discussion of evaluation findings. Moreover, when this criterion was met, we went on to examine *less possible* (??) findings. When the addition of *less possible* to *probable* and *possible* findings would increase the strength of the net effect without altering its direction, our final estimates of net effect included these findings. When *less possible* findings contradicted more probable findings, they were disregarded on the assumption that they would more likely introduce random noise or bias than valid signal. Finally, when the net observed effect for a particular domain was paralleled by a net predicted effect *in the same direction and of roughly the same magnitude*, the net observed effect was not considered "consequential," whatever its magnitude, given the strong likelihood of bias arising from differences in the pretreatment characteristics of PDC and Comparison groups. Since all first-order findings are also reported, the reader may judge for him/herself whether or not the end result of this decision-making process is convincing.

The formats used to present findings from net effect analyses are illustrated with hypothetical data in Figure 11-2 and Table 11-5. The range of net effects observed across sites and in the aggregate sample is represented by a horizontal bar for each outcome domain (here labeled "A" and "B"). Sites where "consequential" effects were found are identified by numerals (1 through 10). The positions of numerals on the horizontal bar indicate magnitudes of net effects and are referenced to the scale (-1.0 through +1.0) at the bottom of the figure. The net effects from aggregate, cross-site tests are indicated by the letter "A" positioned beneath the bar. Vertical dashed lines at -.25 and +.25 indicate our cut-offs; sites whose net effects fall within the -.25 to +.25 interval are not represented by numerals. The table accompanying the figure presents the values of net observed effects derived from analyses of (1) *probable* effects only, (2) *probable* and *possible* effects (Prob/?), and (3) *probable, possible,* and *less possible* effects (P/?/??). It also reports net predicted effects from Design 1. This information is provided for each outcome domain, each site (S1, S2, etc.), and the aggregate sample (Agg). Cells containing "consequential" effects summarized in the figure are highlighted in the table by heavy lines defining their boundaries. Effects columns containing net observed effects of +/- .25 or greater that are counterbalanced by net predicted effects in the same direction and of roughly the same magnitude are highlighted by diagonal lines drawn through the relevant cells.

Based on the hypothetical findings in Figure 11-2 and Table 11-5 we would identify four sites as exhibiting effects of consequence in outcome domain A--three in a positive, PDC-favoring direction (2, 6, and 7); one in a negative, Comparison-favoring direction (4). At one site (9), the net predicted effect was both in the same direction as observed effects and quite large (+.75), casting serious doubt on a conclusion that the observed net positive effect (+.63) was caused by the PDC intervention rather than pre-treatment differences between groups. Predicted effects are used to qualify observed effects only when both net predicted and net observed effects for a domain are +/- .25 or greater and when predicted and observed effects coincide for particular outcome variables within the domain. Although more complicated procedures for taking predictions into account were tried, none offered any obvious advantage over the procedure described above.

Figure II-2

Hypothetical Net Site-Level Effects

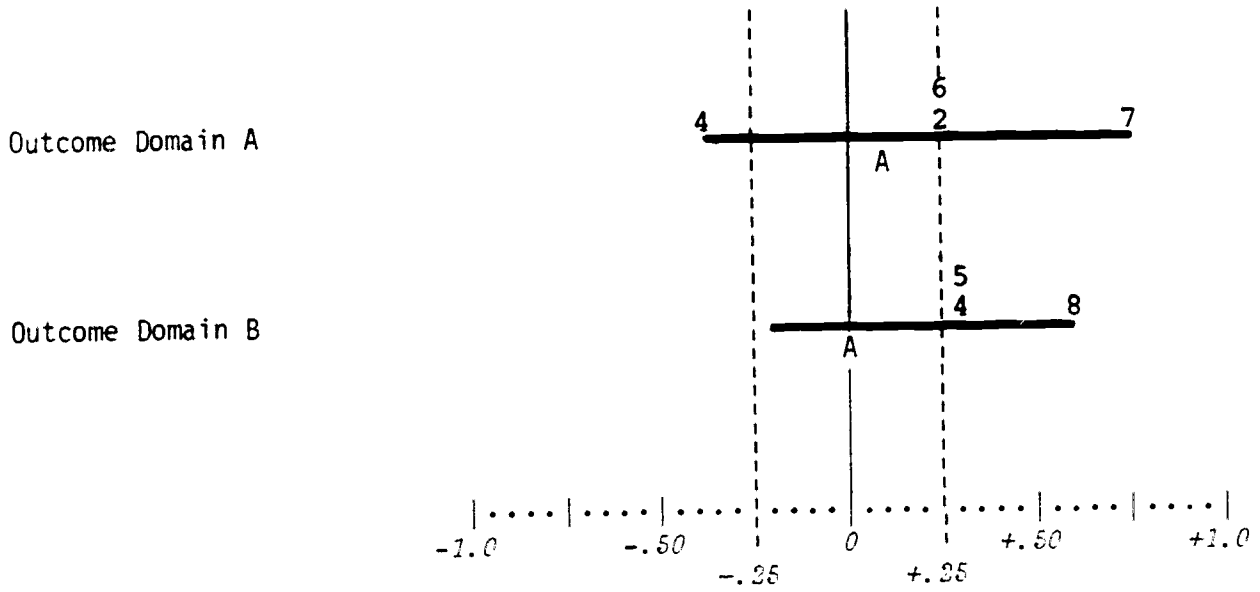


Table II-5

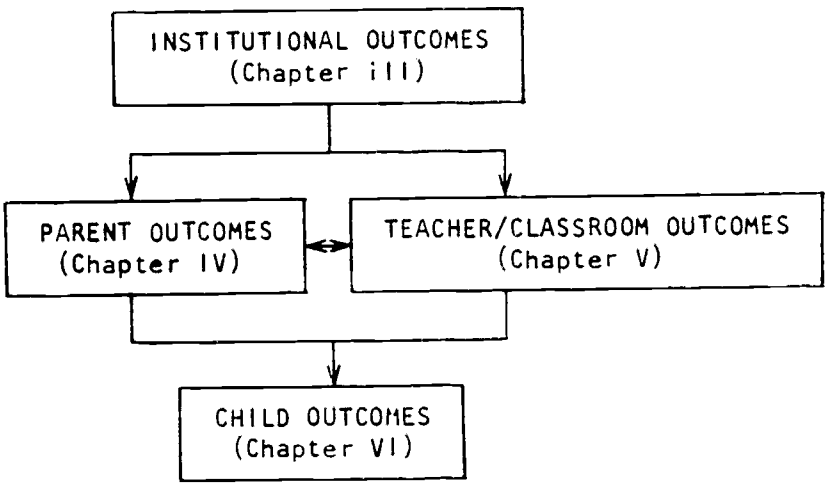
Hypothetical
Summary of Net Effects by Site and Overall
at Three Levels of Probability
and Summary of Net Predictions

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Agg
Outcome Domain A	probable	0	.13	0	-.25	0	.13	.38	-.13	.50	0	0
	prob/?	.13	.25	0	-.25	-.13	.25	.50	-.13	.50	.13	.10
	p/?/??	.13	.25	0	-.38	-.25	.25	.75	-.13	.63	.25	.10
	predict	0	-.13	0	-.13	-.13	0	.13	0	.75	0	0
Outcome Domain B	probable	.10	0	.20	.30	.20	0	-.10	.10	0	0	0
	prob/?	.10	-.10	.20	.30	.30	0	-.10	.30	0	0	0
	p/?/??	.20	-.10	.20	.30	.30	0	-.20	.60	.20	0	0
	predict	0	0	.10	.10	-.20	0	.10	0	0	.10	0

As actual findings are presented in Chapters IV through VI, the reader will be referred back to this section of the report for guidance in interpreting figures and tables.

Synthesis of Findings Across Levels of the Model

As we move from level to level of the intervention model--

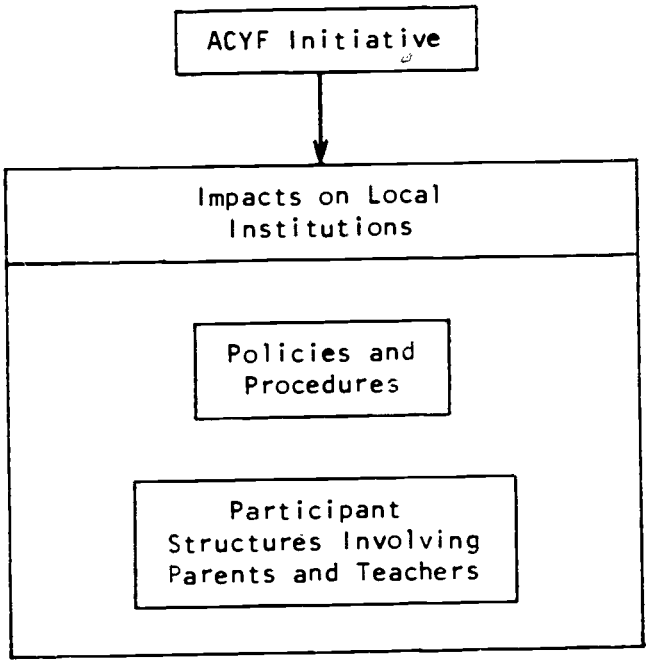


--an effort is made to generate expectations for succeeding levels and to evaluate findings in relation to preceding levels. Although the research design of the longitudinal evaluation severely limits our ability to mathematically model and test relations among institutional, parent, teacher/classroom, and child outcomes, logical analysis of these relations significantly shapes our final conclusions about PDC's impacts (Chapter VII) and how they were or were not achieved.

III

PDC'S INFLUENCE ON LOCAL INSTITUTIONS

According to ACYF's intervention model the PDC initiative would achieve its ultimate objective of enhancing children's social competence by first altering the institutional features of participating Head Start centers and public elementary schools (cf., Figure II-1, in Chapter II.



It was at the institutional level that ACYF intervened, holding local projects accountable for implementing Guidelines that prescribed a common set of institutional features required of all participating Head Start centers and schools. Though it was assumed that the ultimate success of the intervention would depend upon changing the behavior of parents and teachers toward children, PDC Guidelines did not require or operationally define specific objectives for the way parents and teachers interacted with children but left the responsibility for innovation at that level in the hands of local participants. Only if prescribed institutional features were implemented were parents and teachers, then children, expected to change.

The evaluation of PDC's influence on local institutions was guided by two general research questions:

1a. Did Head Start centers and public elementary schools participating in PDC implement the institutional features prescribed in the PDC Guidelines?

1b. Did implementation of PDC Guidelines produce institutional features in PDC centers and schools that were different from those present in non-PDC centers and schools within the same community?

The first of these questions (1a) was elaborated into two, more specific, questions--one concerned with minimal Guideline compliance; the other, with level of Guideline implementation:

- Compliance. Did PDC centers and schools achieve and maintain minimally acceptable levels of model implementation as defined by "required elements" specified in the Guidelines? (i.e., Did PDC projects satisfy the conditions for continuation of their grants?)

- Level of implementation. To what extent did PDC centers and schools go beyond the "letter" (minimal requirements) of the Guidelines to implement their "intent" as implied by ACYF's discussion of goals and basic principles underlying PDC?

The second question (1b) regarding PDC/Comparison difference was asked of both Guideline compliance and level of Guideline implementation:

- Was mere compliance with the letter of the Guidelines sufficient to differentiate PDC from Comparison institutions?

- Did PDC institutions exhibit higher levels of implementation of features prescribed or implied in the Guidelines than Comparison institutions in the same community?

Unless implementation of prescribed/implied institutional features differentiated PDC from Comparison centers and schools in the desired direction in at least some sites, there would be no reason to expect differences in teacher, parent, or child outcomes that could be attributed to the intended intervention rather than to pre-existing differences between groups or to other factors operating in the environment.

As previously mentioned in Chapter I, early indications of site-level variation in PDC programs led us to focus the evaluation on site-level effects of ACYF's intervention. Though analyses of aggregate, cross-site effects offered much greater statistical power, we believed that the results of such analyses would be largely uninterpretable, perhaps under any circumstances, but certainly so if not guided by a thorough understanding of program implementation and outcomes at each site. In keeping with the overall design of the evaluation, analyses of Guideline implementation were conducted and are reported at the site level.

In describing the patterns of implementation of PDC institutional features in this chapter, no systematic effort has been made to explain the patterns observed at particular sites. The reader who is interested in a fuller understanding of the process through which particular patterns of implementation emerged and changed over time is referred to Volume II of this report--The Process of Program Implementation in PDC.

The first section of this chapter--Measurement of Guideline Implementation--identifies the sources of data available to the evaluation, defines each Guideline requirement for which implementation was assessed, and describes the procedures used to aggregate implementation ratings for purposes of characterizing implementation at each site. The second section--Implementation Findings--describes Guideline-relevant features of PDC and Comparison institutions at each site, assessing Guideline compliance, level of implementation, and degree of PDC/Comparison difference.

MEASUREMENT/ANALYSIS OF GUIDELINE IMPLEMENTATION

Before presenting the findings from analyses of institutional features, we shall describe the basic methodology of this facet of the longitudinal evaluation. The discussion that follows is necessarily somewhat tedious, and the reader may wish to skip to the next section--Institutional Findings--returning to this section only as necessary to ascertain the operational meaning of particular variables or to evaluate specific data reduction procedures.

Sources of Data

Information about the implementation of institutional features prescribed by the PDC Guidelines was drawn from diverse sources, and data availability was somewhat uneven across time and sites as indicated in Table III-1.

IRI Implementation Ratings

Head Start implementation data were collected in the spring of 1977 by site-visit teams using the Implementation Rating Instrument (IRI: Smith, Love, Morris, Spencer, Ispa, & Rosario, 1977) to evaluate level of Guideline implementation. Because it was not possible to secure U.S. Office of Management and Budget clearance for the interview schedules in time for data collection (an acute problem for most federally-funded research during the late seventies), data could only be collected on a "pilot" basis at nine sites, interviewing no more than one person at each site with each specialized interview schedule. No data could be collected from Comparison schools, teachers, or parents. Of the nine sites where data were collected, eight were among the ten sites that remained in the longitudinal evaluation.

Table III-1

Institutional Data Sources

Data Source	Cohort Grade and School Year				
	Head Start 1976-77	Kindergarten 1977-78	First 1978-79	Second 1979-80	Third 1980-81
Implementation Ratings (IRI)	PDC at 8 sites				
Pacific Consultants' Case Studies		No implementation data collected	PDC at all sites		
Site Visit Interviews			PDC and selected Comp schools at all sites	PDC and all Comp schools at all sites	PDC at all sites
Teacher Interviews			all P and C cohort teachers and selected others at all sites	all P and C cohort teachers and selected others at all sites	all P and C cohort teachers and selected others at all sites
Parent Interviews			all P and C cohort parents at all sites	all P and C cohort parents at all sites	all P and C cohort parents at all sites

The raw data consisted of responses to semi-structured interviews conducted with PDC program personnel and parents as well as written records of program policies, plans, and activities. The IRI provided a systematic approach to rating the implementation level of each operational guideline on a four-point scale:

1 = no evidence of Guideline implementation
2 = low level of implementation
3 = moderate level of implementation
4 = high level of implementation

In the analyses reported in this chapter, IRI ratings were rescaled as follows:

- Compliance

  Not in compliance = 1
  At least minimal compliance = 2-4

- Level of implementation

  Low = 1 and 2
  Moderate = 3
  High = 4

The IRI methodology and findings have been described in detail in an earlier report (Smith et al., 1977).

Site-Visit Interviews and Pacific Consultants' Case Studies

Following a scheduled (by contract) hiatus of one year (1977-78) in the collection of implementation data, site-visit interviews resumed. In the spring of 1979, the technical assistance contractor, Pacific Consultants, sent staff to all sites to collect data by interview and from written records for comprehensive case studies of each PDC project. During that same period, High/Scope staff also visited all sites and interviewed PDC coordinators, school district administrators, Community Action Program directors, all PDC and selected Comparison school principals, Comparison support services staff, Head Start directors, and parents. In the spring of 1980, High/Scope staff again visited all sites and interviewed PDC coordinators, PDC staff responsible for implementing each Guideline component, school district administrators, Head Start directors, and all PDC and Comparison principals. A final visit by High/Scope staff was made in the spring of 1981. At that time, interviews were conducted with PDC teachers and aides, PDC parents, Head Start directors, district administrators, and other knowledgeable individuals associated with the program. Data from all of these interviews were available both in the form of transcribed, often paraphrased, responses to interview questions and in the form of site-visit reports, which integrated and interpreted the interview data. Pacific Consultants' case study data (Pacific Consultants, 1979) had been largely synthesized into more general descriptive statements. The content of the various interviews conducted by High/Scope staff is outlined for each category of respondent and each year in Appendix C.

Ratings of implementation based on interview and case study data were accomplished by a team of researchers who had been responsible for the original data collection. First, the team attempted to develop operational definitions of each PDC Guideline requirement, building upon prior work (e.g., Smith et al., 1977). Requirements that were not already more or less operationally defined in the published Guidelines were excluded from further consideration since we did not wish to impose our view of the world on the definition of requirements after the fact. Because no projects had been held narrowly accountable for the more ambiguous requirements anyway, this decision seemed reasonable.

Next, each team member undertook a comprehensive review of all qualitative data pertaining to one or more sites with which s/he was directly acquainted. Information relevant to evaluating the implementation of requirements in all component areas was summarized by program, school, and year on special data reduction forms, and preliminary ratings were made of compliance and of program difference in compliance. Following this initial summarization and rating, the person supervising the implementation rating effort examined all site-level compliance ratings in relation to data summaries, identifying areas of disagreement among raters regarding the operational meaning of compliance for particular requirements. Disagreements were resolved through discussion, and ratings were revised as necessary.

Ratings of implementation level were made by the supervisor for all sites, utilizing data summaries and compliance ratings, consulting with site visitors to clarify problematic situations, reviewing preliminary ratings of implementation level with site visitors, and revising ratings as necessary to reflect both the unique knowledge of site visitors about each site and the consensus among team members about the operational meaning of implementation levels. All ratings were the product of both the responsible team member and the supervisor, with an effort made to maintain consensus among all team members as to rating rules and their application. Consequently, the rating process was inter-subjective, if not objective, and produced ratings that we believe to have been replicable at least among members of the team. We did not, however, obtain quantitative estimates of the reliability of ratings.

Since the site-visit interviews from 1977 through 1981 were not specifically designed to monitor all institutional features of PDC and Comparison centers and schools in each project year, there was substantial missing data in our records. When relevant information was truly unknown, no interpolation or extrapolation was undertaken to complete the record. However, team members were encouraged to draw upon their personal knowledge of particular sites (knowledge that had, in some cases, built up over a period of seven years of first-hand contact) to supplement written records.

Another problem impeding data reduction was within-site, within-program variance resulting from differences in Guideline implementation among the several schools/centers constituting PDC or Comparison at particular sites. When this problem was encountered, program ratings were made to reflect the "central tendency" among however many schools/centers were involved. This

estimation process was essentially qualitative and produced site-level program ratings that, under conditions of grossly differential implementation among schools, may have little practical meaning. However, since it was impractical to conduct the full evaluation at the school or center level, there was no alternative but to generate some characterization of the "average" school/center representing PDC and Comparison at each site.

Site visit and case study data were used to generate annual (1979-1981) compliance and implementation ratings of all but three Guideline requirements (having to do with teacher training) for both PDC and Comparison programs at each site. The criteria used to assess compliance and level of implementation for each requirement (Guideline subcomponent) are briefly described in Table II-2 (following).

Teacher Interview

The Teacher Interview developed for the PDC evaluation comprised an invariant sequence of standard questions and probes, some of which structured responses into a priori categories, while others permitted open-ended answers. The interview schedule changed somewhat from the spring of 1979, when it was first used, to spring 1981. The only significant change with respect to items measuring Guideline implementation involved the addition of items measuring teachers' level of participation in different sorts of training in 1980 and 1981. Interview items considered in this chapter are reproduced in Appendix C.

An effort was made to interview all PDC and Comparison teachers of children in the evaluation cohort. In addition, one teacher was randomly selected from each program grade level (K-3) not represented in the sample of cohort teachers at each PDC and matched Comparison school, and from each feeder Head Start center. Data from all interviewed teachers were considered in evaluating program implementation. Although Teacher Interview data were already in ordinal scale form, it was necessary to transform these scales into two-point (yes/no) ratings of compliance and three-point (low/mid/high) ratings of implementation in order to aggregate these data with data from other instruments. These scale transformations are specified in Appendix C together with actual interview questions and original scales.

Parent Interview

The Parent Interview developed for the PDC evaluation was similar in format but not content to the Teacher Interview described above. Although the interview schedule changed somewhat from year to year, the single item (number of parents working in school as volunteers and paid aides) analyzed in this chapter was constant from 1979 through 1981. Since the Parent Interview was only administered to parents of children in the evaluation cohort, the picture these data provide of parent participation in school may not be representative of the much larger number of parents of children attending PDC and Comparison schools. Thus, information from the Parent Interview was used only to confirm or qualify ratings based on site-visit interview data.

Guideline Requirements: Component Areas and Subcomponents

The PDC Guidelines were organized into six major program areas, referred to as components: Administration, Education, Bilingual/Bicultural/Multicultural Education, Education of Handicapped Children, Parent Involvement, and Developmental Support Services.[1] Each component was defined in terms of specific program subcomponents. In assessing Guideline implementation we considered all subcomponents that were sufficiently well defined either in ACYF documents or in mutual understandings between ACYF and local sites to permit meaningful and reliable rating. The thirty-five subcomponents that were ultimately rated are described in Table III-2 along with brief statements of criteria used to assess compliance and implementation level. Data sources are also indicated.

Procedures for Aggregating Ratings

The result of preliminary data reduction efforts was a set of annual implementation ratings of each program subcomponent for both PDC and Comparison at each site. No ratings were made for 1977-78, when site-visit interviews were temporarily suspended; no ratings were made for 1976-77 at two sites (2 and 9) not included in the IRI study; and it was occasionally impossible to rate the implementation of particular program subcomponents at particular sites due to the unavailability of information in either written records or human memory.

All available ratings were tabled in subcomponent-by-year matrices for each site and rating type (compliance/level). Rather than create parallel sets of matrices for each program, we produced compliance (yes/no) and implementation level (low/mid/high) matrices for PDC and corresponding matrices of PDC - Comparison difference in compliance and level, which contained "+" for PDC-favoring differences, "=" for no difference, and "-" for Comparison-favoring differences. In 1979-80 and 1980-81, multiple ratings of the same subcomponent were sometimes available from Teacher Interviews

[1] A seventh component area--Training--was identified in published Guidelines. For purposes of this evaluation the training requirements have been merged into other component areas with which they were functionally associated.

Table III-2

Guideline Requirements: Component Areas and Program Subcomponents

COMPONENT 1: ADMINISTRATION

Subcomponent A: Staffing. Each PDC project was expected to hire the following staff: a full-time PDC coordinator; a part-time (at least 50%) support services coordinator; a part-time (no fraction specified) parent involvement coordinator.

Compliance. Minimum staffing specified above or, in the case of Comparison schools, its functional equivalent.

Level. Higher levels attained by having additional paid staff or paid staff time allocated to program functions specified in Guidelines (component areas 2-6, below).

Data source. Site visit interviews, case studies, IRI.

Subcomponent B: Component Responsibility. Each PDC project was expected to assign specific responsibility for the Education, Bilingual/Bicultural, Handicapped, and Training components to members of school/center staff.

Compliance. Evidence of responsibilities having been formally assigned at the school/center or project level. Having persons responsible for these functions at the district level was not sufficient for compliance.

Level. Level of implementation was determined by the level of activity of those responsible for specified functions.

Data source. Site visit interviews, case studies, IRI.

Subcomponent C: PDC Council Responsibility/Authority. Each PDC project was expected to form a PDC Council that would determine the allocation of project resources, ensure the implementation of all required program components, and maintain close coordination of Head Start and elementary school programs.

Compliance. Evidence that such a Council existed and met at least occasionally. Analogous advisory councils were identified for many Comparison schools and their responsibility/authority was compared with PDC criteria.

Level. Level of implementation was determined by the level of activity of Councils in carrying out their specified responsibilities.

Data source. Site visit interviews, IRI.

Subcomponent D: Broad Representation on PDC Council. Each PDC Council was expected to include representatives from the following groups: parents of PDC Head Start and elementary school children; members of the Head Start Policy Council and local school board; Head Start and school administrators; Head Start and school staff; and local community health and social service agencies.

Table III-2, continued

Compliance. Evidence that representatives from each group specified above had been appointed to the advisory council (PDC Council or its analogue in Comparison schools/centers).
Level. Level of implementation was determined by the level of involvement (attendance, initiative) of representatives from these diverse groups.
Data source. Site visit interviews, case studies, IRI.

Subcomponent E: Council Communication with Parents. Provision was to be made for regular communication by the Council to parents of information about program activities and project resource allocation.
Compliance. Evidence of mechanisms for communication (e.g., newsletters, workshops, meetings) and at least occasional acts of communication.
Level. Level of implementation was determined by frequency and effectiveness of communication.
Data source. Site visit interviews, IRI.

Subcomponent F: Council Communication with School/Center Staff. As for E, above.

Subcomponent G: Training for Parents in Policy- and Decision-Making Skills. PDC projects were expected to provide training in policy- and decision-making skills to all interested parents, but particularly to parents on the PDC Council and its subcommittees.
Compliance. Any evidence of training in this area for parents.
Level. Implementation level was determined by level of activity and size of parent audience reached.
Data source. Site visit interviews, IRI.

Subcomponent H: Training for Parents and Staff in Goals and Requirements of PDC. All program staff, Council members, and parents were to receive training in the general concept of developmental continuity, the basic principles and requirements of the PDC Guidelines, and the local project's coordinated Head Start through early elementary program.
Compliance. Any evidence of training in this area. For Comparison schools, any evidence of analogous training in the goals and methods of new or unusual school or district initiatives.
Level. Level of implementation determined by level of training activity and diversity/size of audience.
Data source. Teacher Interview, site visit interviews, IRI.

Table III-2, continued

COMPONENT 2: EDUCATION

Subcomponent A: Coordinated Curriculum from Head Start through Early Elementary Years. Each PDC project was expected to develop a coordinated curriculum that specified learning objectives and developmental landmarks from Head Start through third grade.

Compliance. Written evidence of such a curriculum.

Level. Implementation level was determined by the level of curriculum development (detail, scope, instructional materials, assessment/monitoring procedures, etc.) and the degree to which the curriculum appeared to be enacted by teachers.

Data source. Site visit interviews, case studies, IRI.

Subcomponent B: Internal Assessment Procedures. Each project was expected to establish procedures for ongoing internal assessment of the educational program that were either formal or informal and that involved staff, parents, and Council members.

Compliance. Any evidence of internal assessment.

Level. Implementation level was determined by level of activity and scope of involvement.

Data source. Site visit interviews.

Subcomponent C: Curriculum Refinement Procedures. Each project was to establish procedures that involved teachers, aides, administrators, resource staff, and parents in ongoing discussion and refinement of the curriculum.

Compliance. Any evidence that such discussions occurred.

Level. Level of implementation was determined by the level of discussion/refinement activity reported by teaching staff.

Data source. Teacher Interview, site visit interviews, IRI.

Subcomponent D: Diagnostic-Evaluative System. Each project was expected to develop/select and implement a formal "diagnostic-evaluative system" for determining the educational needs of individual children.

Compliance. Evidence that a formal evaluative system, with at least diagnostic potential, was being implemented.

Level. Level of implementation was determined by the scope of the evaluative system and the degree to which it was put to regular, diagnostic-prescriptive use.

Data source. Site visit interviews.

Subcomponent E: Teacher Training for Individualizing Instruction. Each project was to provide all Head Start and elementary school teachers with training in methods for individualizing instruction.

Table III-2, continued

Compliance. Any evidence of such training.
Level. Implementation level determined by frequency of such training.
Data source. Teacher Interview.

Subcomponent F: Training for Teachers in Child Growth and Development. As for E, above.

Subcomponent G: Training for Teachers and Parents in Preventive Health, First Aid, and Safety. As for E, above.

COMPONENT 3: BILINGUAL/BICULTURAL/MULTICULTURAL EDUCATION

Subcomponent A: Coordinated Head Start-Elementary School Approach. Head Start and elementary school staff were to coordinate services to bilingual/bicultural children and families and to develop a coordinated approach to multicultural education for all children.
Compliance. Any evidence of center-school coordination in these areas.
Level. Implementation level determined by level of activity and effectiveness of coordination.
Data source. Site visit interviews, IRI.

Subcomponent B: Teacher Training in Bilingual/Bicultural/Multicultural Education Needs and Methods. Training was to be provided to all Head Start and elementary school teachers in the special needs (language, self-concept, cultural) of bilingual/bicultural children and in methods for bilingual/bicultural/multicultural education.
Compliance. Any evidence of such training.
Level. Implementation level determined by frequency of such training.
Data source. Teacher Interview, IRI, site visit interviews.

COMPONENT 4: EDUCATION OF HANDICAPPED CHILDREN

Subcomponent A: Coordinated Program. Head Start and elementary schools were expected to systematically coordinate services to handicapped children.
Compliance. Any evidence of systematic coordination.
Level. Implementation level determined by level of activity and effectiveness of coordination.
Data source. Site visit interviews, IRI.

Subcomponent B: Early Diagnosis. Provisions were to be made for the early diagnosis of learning disabilities.

Table III-2, continued

Compliance. Any evidence of regular screening beginning in Head Start.
Level. Level of implementation determined by regularity of screening and the extent to which systematic prescriptions were made and implemented.
Data source. Site visit interviews, IRI.

Subcomponent C: Provision of Special Materials and Structural Modifications. Each project was expected to create physical environments in centers and schools that could accommodate physically handicapped children.
Compliance. Any evidence of special provisions for the physically handicapped.
Level. Implementation level determined by the extent of special provisions.
Data source. Site visit interviews, informal observations, IRI.

Subcomponent D: Mainstreaming. Each project was expected to mainstream handicapped children in regular classrooms to the maximum extent possible from Head Start through third grade.
Compliance. Any evidence of mainstreaming.
Level. Implementation level determined by the extent of mainstreaming.
Data source. Site visit interviews, IRI.

Subcomponent E: Annual Survey of Handicapped Children. Each project was expected to conduct an annual survey of handicapped children, reviewing individual progress and needs and planning future services.
Compliance. Any evidence of annual survey.
Level. Level of implementation determined by scope and thoroughness of review and planning process.
Data source. Site visit interviews, IRI.

Subcomponent F: Staff Training in Needs of Handicapped Children and Educational Methods. Teachers, aides, and classroom volunteers were to receive training in the special needs of handicapped children and in how to meet those needs within the classroom setting.
Compliance. Any evidence of such training.
Level. Implementation level was determined by frequency of such training as reported by teachers.
Data source. Teacher Interview, site visit interviews, IRI.

Table III-2, continued

COMPONENT 5: PARENT INVOLVEMENT

Subcomponent A: Coordinated Parent Involvement Program. Head Start and elementary schools were expected to develop coordinated parent involvement programs that would extend the high levels of parent involvement typically achieved in Head Start into the elementary school.

Compliance. Any evidence of a coordinated approach.
Level. Implementation level determined by level of coordination as evident in staffing patterns, lines of communication, and specific activities.
Data source. Site visit interviews, IRI.

Subcomponent B: Use of Parents as Resource Persons in Classroom. Head Start and elementary schools were expected to utilize parents as resource persons in the classroom.

Compliance. Evidence that any parents made educational contributions in classrooms.
Level. Implementation level was determined by the proportion of visiting parents who made educational contributions.
Data source. Teacher Interview, site visit interviews, IRI.

Subcomponent C: Parents as Volunteers and Paid Aides. Parents were to be involved in classrooms as volunteers and/or paid aides.

Compliance. Any evidence of such formal involvement.
Level. Level of implementation determined by proportion of parents who worked in program classrooms as volunteers or paid aides.
Data source. Parent Interview, site visit interviews.

Subcomponent D: Parent Training in Child Growth and Development. Parents were to be offered training in child growth and development that would enable them to better understand and meet their children's changing needs.

Compliance. Any evidence of such training
Level. Implementation level determined by level of training activity.
Data source. Site visit interviews.

Subcomponent E: Training for Parents in How to Support the Education of Their Children Outside of School. As for D, above.

Subcomponent F: Training for Parents and Teachers in Working with Each Other. As for D, above.

Data source. Teacher Interview, site visit interviews, IRI.

Table III-2, continued

COMPONENT 6: DEVELOPMENTAL SUPPORT SERVICES

Subcomponent A: Coordinated Support Services Programming. Each project was expected to develop a coordinated approach to providing support services (i.e., nutritional, medical, dental, mental health, and social services) that would extend the comprehensive Head Start program into the elementary school.

Compliance. Any evidence of coordinated services from Head Start through third grade.

Level. Implementation level determined by the comprehensiveness of services provided in kindergarten through third grade.

Data source. Site visit interviews, IRI.

Subcomponent B: Initial Screening of Children. The nutritional, medical, dental, mental health, and social service needs of all enrolling children (HS-3) were to be assessed regardless of age or grade level.

Compliance. Any evidence of systematic screening.

Level. Implementation level determined by the thoroughness of screening.

Data source. Site visit interviews, IRI.

Subcomponent C: Mealtime and Snacktimes as Opportunities for Learning. Each project was expected to devise ways of utilizing mealtimes and snacktimes for learning (particularly health and nutrition education) rather than merely eating. This was already a feature of Head Start programs.

Compliance. Any evidence that mealtimes/snacktimes were systematically used for learning.

Level. Implementation level was determined by the degree to which the learning potential of mealtimes/snacktimes was exploited.

Data source. Site visit interviews.

Subcomponent D: Communication of Child Health Information to Parents. In an effort to involve parents in the health care of their children, PDC programs were expected to provide parents with information about physical examinations, follow-up treatments, and immunizations of their children.

Compliance. Any evidence of such communication.

Level. Implementation level determined by the regularity and thoroughness of such communication.

Data source. Site visit interviews, IRI.

Table III-2, continued

Subcomponent E: Familiarization of Parents with Available Health Resources. Parents were to be provided with information about all health resources available in their communities.

Compliance. Any evidence of such communication.

Level. Implementation level determined by thoroughness of such communication and by effectiveness of dissemination mechanisms.

Data source. Site visit interviews, IRI.

Subcomponent F: Continuity of Record-Keeping. Each project was expected to establish a continuous and comprehensive system for maintaining children's medical and dental records from Head Start through grade three.

Compliance. Any evidence of continuous record-keeping.

Level. Level of implementation determined by comprehensiveness of system, update capability, and data accessibility.

Data source. Site visit interviews.

and site-visit interviews (see Table III-2). When this occurred, ratings of Teacher Interview data were used unless strongly contradicted by site-visit interview data and first-hand experience.

Once rating matrices had been constructed, ratings were aggregated to subcomponent, component, and overall levels across all four measurement years (1977, 1979-1981) to characterize PDC compliance and level of implementation as well as PDC - Comparison difference at each site. The following aggregation procedures were employed:

- Subcomponents were characterized by the median value of annual ratings;
- Components were characterized by the median value of all annual ratings of all associated subcomponents;
- Overall implementation at each site was characterized by the median value of all component ratings, thereby weighting components with different numbers of subcomponents equally.

We selected the median as our measure of central tendency primarily to avoid introducing a false sense of precision by computing mean values for rather humble ordinal scale variables (yes/no and high/mid/low). When the number of cases is even, convention has it that the median value is the arithmetic mean of the two middle cases; however, in keeping with our decision to treat these ratings as simple ordinal scales, we applied the following rules in assigning median values when confronted with an even number of cases:

- When the two middle cases had equal values, the median was given that value;
- When the two middle cases had different values, we represented the hypothetical median by both values--Y/N, H/M, M/L, +/=, =/-, and infrequently H/L or +/-.

Though this solution is not elegant, it seems appropriate given the available data.

In evaluating implementation we have tried to consider consistency as well as characteristic (i.e., median) level. Local PDC programs operated for five years in a demonstration mode, during which time the cohort of children included in the longitudinal evaluation moved from Head Start through third grade. Over this period, there was substantial opportunity for variation in implementation level as program staff turned over, local education policies shifted, the purchasing power of PDC grants declined, and so forth. In addition, there were indications that local sites deliberately emphasized implementation of some subcomponents at the expense of others, introducing variation in implementation level across subcomponents/components at particular points in time. In order to reflect the underlying

consistency or inconsistency of aggregate ratings, we estimated the degree of dispersion of scores around each aggregate (i.e., median) rating. Dispersion was determined to be either "high" or "low" according to the following rules:

- "High" if less than 2/3 of ratings fell within the category containing the median; "low" otherwise.

or

- In the case of implementation level ratings, "high" if there were (a) both high and low ratings but no moderate ratings and (b) less than 75% of ratings fell within the category containing the median; "low" otherwise.

In situations where the hypothetical median fell between categories (e.g., H/M, H/M), dispersion was always high. In the next section--Implementation Findings--high dispersion is indicated by an asterisk associated with reported median values (e.g., M*). Since overall ratings represented median values of component not subcomponent ratings, estimates of dispersion were based upon variance among component ratings disregarding variability of the annual subcomponent ratings, which underlay component ratings. Although the evaluation design does not permit mathematical modeling of subtle relationships between variations in implementation and outcomes, information about variation in implementation across time and program areas can shape one's expectations regarding outcomes as well as one's interpretation of whatever outcomes are found.

Adequacy of Measurement

We certainly attempted to measure implementation of all facets of the PDC model that had been operationally defined by ACYF. How well measurement of these institutional features was accomplished must remain a matter of opinion. The strongest evidence of the validity of aggregate ratings that we can present comes from former site visitors who generally confirm the component-level and overall characterizations of PDC implementation levels and PDC/Comparison difference reported in the next section.

As was explained in Chapter I, ACYF's Guidelines were intended to define a common, cross-site program model in terms of institutional/organizational features. At the same time, ACYF expected each local project to develop fully operational interventions of their own, working within the framework provided by the Guidelines, but going substantially beyond required institutional features to involve teachers and parents in new relationships with one another and with children. Whether or not distinctive local elaborations of the program were developed and implemented will be considered in Chapters IV and V where we examine parent and teacher/classroom outcomes. The measurement of implementation of locally elaborated programs was considerably more problematic than measuring implementation of the basic model framework provided by ACYF, though both tasks proved difficult and were imperfectly accomplished.

IMPLEMENTATION FINDINGS

The discussion that follows is concerned primarily with evaluating the implementation of major program components defined by clusters of PDC Guideline requirements, or subcomponents. The six component areas considered in this evaluation are:

- Administration
- Education
- Bilingual/Bicultural/Multicultural Education
- Education of Handicapped Children
- Parent Involvement
- Developmental Support Services

Subcomponents and associated rating criteria have already been defined in Table III-2 (above). All subcomponent ratings are presented by site in Appendix D for perusal by the interested reader; they will be considered here only as necessary to elucidate component-level and overall findings. Findings are first presented and discussed for Guideline compliance, then for level of Guideline implementation.

Compliance with PDC Guideline Requirements

All PDC projects were expected, at a minimum, to comply with Guideline requirements. Compliance could be achieved with only "nominal implementation." For example, Subcomponent B of Component 3--Bilingual/Bicultural/Multicultural Education--required training of Head Start and elementary school teachers in meeting the special needs of bilingual/bicultural children and/or furthering the multicultural education of all children. If during the course of each school year some such inservice training session was offered (by the district if not the project) and attended by someone teaching in the program, a project was judged to have complied with this Guideline requirement. To achieve compliance it was not necessary that all teachers be trained, that training be effective, or that classroom instruction accomplish the goals of bilingual/bicultural/multicultural education.

PDC Compliance

Did PDC centers and schools achieve and maintain minimally acceptable levels of model implementation as defined by required elements specified in the Guidelines?

Component and overall implementation ratings presented in Table III-3 indicate that all local projects tended to be in compliance with Guideline requirements. For only one component were there indications of serious non-compliance at particular sites. At sites 1, 3, and 4, it appears that Component 3--Bilingual/Bicultural/Multicultural Education--was not consistently implemented; all three sites received "Y/N" aggregate ratings. Site 9 was judged to have been generally out of compliance with Component 3 requirements, receiving an aggregate rating of "N." In all four cases, inspection of the matrices of annual ratings (Appendix D) revealed that non-compliance resulted primarily from the absence of a "coordinated approach" to bilingual/bicultural/multicultural education in all or most project years.

Though the central tendency of compliance ratings would suggest few serious problems of consistent non-compliance, in eight instances the annual ratings underlying "Y" component ratings exhibited high dispersion (*) and in only six instances were all subcomponents rated "Y" (†) in all years. Interestingly four of these six instances of no variance among annual ratings occurred for Component 4--Education of Handicapped Children--a program area in which implementation was largely determined by local, state, and federal regulations for implementing Public Law 94-142 (the Education for All Handicapped Children Act). Although the law, passed by the U.S. Congress in 1975, did not take effect nationally until 1978, several states in which PDC sites were located implemented their own versions prior to that date.

PDC-Comparison Difference in Compliance

Was mere compliance with the letter of the Guidelines sufficient to differentiate PDC from Comparison institutions?

Findings presented in Table III-4 for overall implementation suggest that mere compliance was not sufficient to differentiate PDC from Comparison schools and centers. The basic ideas embodied in the PDC concept were not unique. Progressive educators across the country would have agreed with PDC's general intent and with many of its methods. Other federally funded compensatory and bilingual/bicultural programs operating in PDC communities incorporated similar guidelines in many areas and provided resources that were to be allocated in similar ways. And the homogenizing effect of P.L. 94-142 on services for handicapped children has already been noted.

Table III-3

PDC Project Compliance with Guidelines

Site	Component Area						Overall Rating
	Administration	Education	Bilingual/Multicultural/ Multicultural Education	Education of Handicapped Children	Parent Involvement	Developmental Support Services	
Site 1	Y	Y	Y/N*	Y	Y*	Y	Y
Site 2	Y	Y	Y	Y	Y	Y	Y†
Site 3	Y	Y	Y/N*	Y†	Y	Y*	Y
Site 4	Y	Y	Y/N*	Y	Y*	Y	Y
Site 5	Y	Y	Y*	Y	Y	Y	Y†
Site 6	Y	Y	Y	Y	Y	Y	Y†
Site 7	Y	Y	Y	Y†	Y	Y†	Y†
Site 8	Y*	Y	Y	Y	Y*	Y	Y†
Site 9	Y	Y	N	Y†	Y	Y*	Y
Site 10	Y	Y	Y*	Y†	Y	Y†	Y†

* Indicates high variance in underlying ratings.
† Indicates no variance in underlying ratings.

Table III-4

PDC-Comparison Difference in Guideline Compliance

Site	Component Area						Overall Rating
	Administration	Education	Bilingual/Bicultural/ Multicultural Education	Education of Handicapped Children	Parent Involvement	Developmental Suport Services	
Site 1	=*	=	=	=	=*	+*	=
Site 2	=*	=*	+/=*	=	+*	=*	=
Site 3	+*	=	=	=	=*	=	=
Site 4	+	=	=	=	=	+*	=
Site 5	+*	=	+*	=	=*	=	=
Site 6	=*	=	+*	=	=	+*	=
Site 7	+	=	=*	=	+/=*	+*	=
Site 8	=	=	=	=	=*	=	=
Site 9	+	=	=	=	=	+/=*	=
Site 10	+*	=	=	=	=	=	=

* Indicates high variance in underlying ratings.

But while PDC was not entirely novel, it did incorporate some unique features and did press for more comprehensive implementation of elements common to other programs. PDC's unique features involved bridging Head Start and public elementary school programs by insisting upon coordinated programming in Components 2 through 6 (Subcomponent A in each Component dealt with Head Start-public school coordination), participation by Head Start parents and staff in the PDC Council (analogous to school councils of various sorts attached to Comparison institutions), and extension of the Head Start support services model into the elementary grades. It was primarily implementation of these Head Start-derived elements that accounted for the few PDC-favoring differences in compliance at the component-level.

Level of Implementation of PDC Guidelines

To have achieved minimal compliance with PDC Guideline requirements was clearly not to have implemented the PDC program envisaged by ACYF. Moderate to high levels of implementation entailed going beyond the "letter" of the Guidelines in directions suggested by ACYF's discussion of PDC goals, basic principles, and possible methods. For example, Subcomponent D of Component 2--Education--required each project to implement a "diagnostic-evaluative" system for determining the educational needs of individual children. Given the surge of interest in educational testing during the 1970's, most school districts had already implemented assessment programs that were sufficient to achieve compliance with PDC Guideline requirements. However, high levels of implementation could only be achieved by integrating assessment programs with instructional programs in such a manner that the results of educational testing systematically influenced educational programming for individual children across the entire age/grade range addressed by the program.

Level of Implementation in PDC

To what extent did PDC centers and schools go beyond the "letter" of the Guidelines to implement their intent as implied by ACYF's discussion of the goals and basic principles underlying PDC?

Findings presented in Table III-5 suggest that local PDC projects generally did go somewhat beyond the letter of the Guidelines. Overall implementation ratings were "Moderate" for eight sites and "High/Moderate" for two (sites 8 and 10). However, it must also be noted the local projects generally did not go far beyond what was minimally expected and that levels of implementation tended to be rather inconsistent at both the component and overall levels (*). No particular component stands out as being unusually well implemented across sites; however, Component 3--Bilingual/Bicultural/Multicultural Education--stands out as being poorly implemented at most sites.

Table III-5

PDC Project Guideline Implementation Level[1]

Site	Component Area						Overall Rating
	Administration	Education	Bilingual/Bicultural/ Multicultural Education	Education of Handicapped Children	Parent Involvement	Developmental Support Services	
Site 1	M/L*	M/L*	L*	L	M*	M*	M/L*
Site 2	M*	M*	M/L*	M	M*	M*	M
Site 3	M*	L	L	H	M*	M*	M*
Site 4	M*	M*	L†	M*	M*	L*	M
Site 5	M	M	L	M*	M	M*	M
Site 6	M*	M*	M*	L*	M*	M*	M
Site 7	H*	M*	M*	M*	M*	H	M
Site 8	L	M*	H	H	M*	H	H/M*
Site 9	M	L	L†	H*	M*	M*	M*
Site 10	H	M*	M/L	H	M*	H	H/M*

* Indicates high variance in underlying ratings.
† Indicates no variance in underlying ratings.

[1] H = High
M = Moderate
L = Low

Inconsistency at the component level arose both from variance across subcomponents and from variance across time. Post hoc analyses indicated that at six of the eight sites for which implementation data were available in both 1976-77 and 1980-81 overall implementation ratings declined from spring 1977 to spring 1981. Fully 60% of component ratings declined over this period, while only 15% of component ratings increased. Although it is not our purpose in this volume to explain patterns and changes in patterns of implementation, (see Volume II, The Process of Institutional Change in PDC) it seems worth mentioning several factors that appear to have promoted decay in implementation levels at some sites. Economic factors played important roles at some sites, specifically decreases in the purchasing power of local project grants accompanied by corresponding decreases in staffing and other resources. Project budget problems were exacerbated in some communities by cutbacks in district budgets that limited district capacity to supplement waning PDC resources. Also significant in some communities was the adoption of district-wide curricula and diagnostic-prescriptive systems that supplanted PDC-developed systems by force. Another factor making it difficult to sustain continuous high implementation levels was turn-over in principals, project and teaching staff. Considerable time and resources were required to train new teaching staff, and they could not always be induced to assume PDC-defined roles that they had played no part in creating and may even have found incompatible with their own skills and professional inclinations.

One additional factor influencing implementation should be mentioned. Certainly by the 1979-80 school year everyone involved in PDC was aware that the program would not go on forever in its current state. Only at site 8 had PDC been significantly institutionalized by virtue of the district's having adopted major elements of the PDC program as the district's program. Elsewhere PDC seemed destined to vanish, but for a few traces, as soon as federal funding terminated, and with PDC would vanish specialized staff positions (coordinators of the project, parent involvement, and support services), specialized roles for mainline staff, various material resources, and the "press to implement" created by the Guidelines and such institutional structures as the PDC Councils. During the final demonstration year (1980-81), some specialized staff were forced to direct energies toward securing new jobs. And in an effort to ensure PDC project staff continuous employment, district administrators in some sites began reassigning PDC staff to other positions (on "hard money") as openings occurred, and before the PDC demonstration terminated. Thus, the PDC program was being dismantled, either figuratively or literally, while it was still being evaluated, less than perfect conditions for measuring effects that some believed would be cumulative through the third grade year.

PDC-Comparison Difference in Implementation Level

Did PDC institutions exhibit higher levels of implementation of features prescribed or implied in the Guidelines than comparison institutions in the same community?

Findings presented in Table III-6 reveal a number of interesting PDC-favoring differences in implementation level and no Comparison-favoring differences. Overall implementation was found to be at least marginally higher in PDC than in Comparison institutions at four sites: two sites received aggregate ratings of implementation difference of "+/="; two, ratings of "+" (with low dispersion). At all four of these sites, however, the absolute level of implementation in PDC was only moderate.

As was the case in evaluating compliance difference, there was no indication that levels of implementation of Component 4--Education of Handicapped Children--differed between PDC and Comparison schools/centers. However, at nine of ten sites, Component 1--Administration--was appreciably better implemented in PDC than in Comparison schools; at six of ten sites, Component 5--Parent Involvement--was better implemented in PDC; and Component 6--Developmental Support Services--was better implemented (+) or marginally better implemented (+/=) at six sites.

It is interesting to note that the three components (1,5 and 6) showing the most PDC-favoring effects were the same three components for which specific responsibility was assigned to specialist members of staff paid out of PDC budgets: the PDC Project Coordinator, Parent Involvement Coordinator, and Support Services Coordinator. Very few Comparison schools had analogous specialists on staff, though most had some access to district support staff who performed similar functions. It should also be noted that many subcomponents in these three areas of responsibility could be implemented to moderate or even high levels by the area coordinator working more or less independently of other project or center/school staff. This was most true of the Developmental Support Services component (6) whose major subcomponents--screening, record-keeping, communication of diagnosis and treatment to parents, etc.--were, in fact, implemented quite independently of other staff and with their blessing. This was least true of the Parent Involvement component (5), which required very considerable cooperation from teachers for effective implementation of parent involvement in the classroom. However, even in the case of Parent Involvement one of the most significant contributions to PDC-favoring differences was generally high implementation of Subcomponent A, requiring coordination of Head Start and PDC parent involvement programs. This subcomponent could be well implemented by specialist Head Start and PDC Parent Involvement Coordinators (sometimes one and the same person) independently of other staff by the Head Start Coordinator's sharing names of parents whose children were about to enter elementary school and the PDC Coordinator's making some specific effort to involve parents of former Head Start children. This is not to demean the accomplishments of specialist staff on the PDC payroll, but simply to point out that the areas of greatest difference between features of PDC and Comparison institutions tended to be areas in which new elements and staff were grafted onto existing institutions without requiring fundamental changes in the institutional status quo. The single possible exception to this generalization is the area of parent involvement where there were some suggestions of more fundamental change in pre-existing institutional procedures and structures at several sites.

Table III-6

PDC-Comparison Difference in Guideline Implementation Level

Site	Component Area						Overall Rating
	Administration	Education	Bilingual/Bicultural/ Multicultural/Education	Education of Handicapped Children	Parent Involvement	Developmental Support Services	
Site 1	+*	=*	=*	=	+	+	+/=*
Site 2	+*	=*	+/=*	=	+*	+/=*	+/=*
Site 3	+	=	=	=	=*	= *	=
Site 4	+*	+*	=	=	=*	=*	=
Site 5	+	=*	=	=	+	+/=	=*
Site 6	+*	=*	+	=	+*	+*	+
Site 7	+	+/=*	+	=	+*	+	+
Site 8	=	=*	=	=	=*	=	=
Site 9	+	=	=	=	+*	=*	=
Site 10	+	=*	=*	=	=*	+*	=

* Indicates high variance in underlying ratings.

Conclusions and Implications

The major findings of the implementation study were as follows:

- Local PDC projects generally achieved and maintained minimally acceptable levels of model implementation--i.e., they were in basic compliance with the terms of their grants as defined by basic requirements in the PDC Guidelines.

- However, local Comparison institutions were also judged to have unwittingly complied with PDC Guidelines in most cases, suggesting that many institutional features of the PDC model were already present in local institutions in at least embryonic form or that they were introduced during the period of this evaluation.

- All local PDC projects went beyond mere Guideline compliance to implement at least some components at moderate or high levels.

- However, overall levels of implementation by PDC projects were at best "high to moderate" (2 sites) and typically "moderate" (8 sites).

- PDC projects generally exhibited inconsistent levels of implementation both across components at particular points in time and within components across time.

- Levels of implementation among PDC projects generally declined from the beginning to the end of the demonstration period.

- Overall implementation appeared to be higher for PDC than Comparison institutions at four sites--substantially higher in two sites and marginally higher in two others.

- Three component areas were found to be better implemented in PDC than Comparison institutions at a majority of sites: Administration, Parent Involvement, and Developmental Support Services. In no case was a component implemented less well in PDC than Comparison.

While these findings suggest that ACYF had some measure of success in its efforts to alter the institutional features of Head Start centers and public elementary schools, none of the ten sites included in the longitudinal evaluation achieved consistently high implementation of the PDC

model as it was defined in terms of institutional policies, procedures, and participant structures. Moreover, the two PDC projects that achieved the highest overall levels of implementation (sites 8 and 10) were matched in this accomplishment by local Comparison schools and centers.[1]

Implications of Findings for Data Analytic Strategy

Though ACYF fully expected local elaborations of PDC programs to be distinctive, albeit compatible with basic goals and principles, it was hoped that all local projects would achieve high levels of implementation of all components operationally defined in the PDC Guidelines. As we have seen, this hope was not realized. Instead, even at the level of Guideline implementation, there were as many PDC programs or "treatments," judging from patterns of component implementation, as there were sites. Under this condition there was no alternative but to pursue the evaluation of program impacts through site-level analysis and interpretation of outcomes.

Implications of Findings for the Success of the Intervention and the Evaluation

According to the logic of the intervention model and the evaluation model, the higher the level of model implementation and the more different PDC and Comparison institutions in directions implied by the Guidelines, the more likely and larger would be the expected PDC impacts on parents and home environments, teachers and classrooms, and children. Under the conditions of program implementation described above--characteristically moderate overall levels of implementation by PDC projects, inconsistent implementation over time and across components, and similar institutional features in PDC and Comparison centers/schools in most program areas at most sites--one would not expect to find strong, or perhaps any, systematic program effects, even if the causal assumptions underlying ACYF's intervention model were perfectly sound. Add to these conditions the necessity of site-level analyses, small site-level samples, as well as the probable biasing effects of site-level sample nonequivalence, and the likelihood of identifying PDC program effects was further diminished. A logical question might be, "Why proceed with an evaluation under these circumstances?"

---

[1] In site 8, the similarity of PDC and Comparison institutions resulted from direct intervention by the local school district administration who in 1978 adopted the basic PDC model for use in all elementary schools. In site 10, the similarity of PDC and Comparison institutions was the product of more complex forces. (See Volume II for a further discussion of these issues.)

Reasons for Proceeding with the Evaluation

Aside from the obvious reason that we were compelled by contract to proceed, two other reasons can be cited. First, there was evidence that some Guideline components were better implemented in PDC than Comparison institutions at some sites. In fact, Components 1, 5 and 6--Administration, Parent Involvement, and Developmental Support Services--were better implemented in PDC institutions at a majority of sites. Second, it was entirely possible that ACYF's intervention stimulated the development of important program innovations at the local level, innovations compatible with ACYF's intentions and growing out of the Guidelines, but neither required nor operationally defined in the basic, cross-site model. Such local elaborations were expected to occur primarily in classroom content, structure, and process. Whether such program elaborations actually occurred will be considered in Chapters IV and V; here, we shall restrict our attention to the possible implications of PDC-favoring differences in the implementation of specific Guideline components and subcomponents for outcomes in the parent/home, teacher/classroom, and child domains.

While overall high implementation of the Guidelines was expected to lead to positive effects in all outcome domains, the PDC intervention model as described in public documents generally did not link implementation of specific subcomponents or components with achievement of specific outcomes. And when one compares component and subcomponent definitions with descriptions of available outcome measures in the parent, teacher/classroom, and child domains, there appear to be relatively few direct relationships. Except for the Developmental Support Services component (6), the impacts of institutional change on children were expected to be mediated by largely unspecified changes in parents and teachers. Although it might have been possible to measure health and nutritional outcomes for children that would have been directly affected by implementation of the Support Services component, the child outcome battery did not include such measures.

The situation is somewhat different in the case of parent and selected teacher/classroom outcomes where some implementation-outcome relations can be hypothesized based on logical or commonsense analysis of available measures and ACYF's intentions. Specifically, it seemed reasonable to assume that teachers' efforts to promote parent involvement, actual levels of parent involvement, and out-of-school impacts on parent-child interactions would be greater in PDC programs that achieved higher levels of implementation of the Parent Involvement component of the Guidelines. Moreover, PDC-favoring differences in these outcomes might be expected at sites where institutional features defining the Parent Involvement component were more evident in PDC than in Comparison schools and centers--i.e., sites 1, 2, 5, 6, 7, and 9 (Table III-6). Whether this expectation was warranted will be considered in Chapters IV and V as Parent and Teacher/Classroom findings are presented and discussed.

Conclusion

In sum, implementation of the total configuration of PDC-prescribed institutional features fell considerably short of expectations at all sites, making it impossible to put ACYF's intended PDC model to a full and fair test. However, implementation of some institutional features was sufficiently higher in PDC than Comparison centers, schools that related program impacts on certain parent and teacher outcomes seemed possible in some sites. Whether or not hypothesized program effects actually materialized will be considered in Chapter IV (Parent Outcomes) and Chapter V (Teacher/Classroom Outcomes), where we also look for evidence of distinctive local program models in data describing parent participation, classroom content, structure, and process.

IV

PARENT OUTCOMES

In order to achieve its ultimate objective of enhancing children's social competence, ACYF formulated an intervention strategy requiring implementation of common institutional/organizational features among participating Head Start centers and public elementary schools at all sites. Within the institutional framework created by implementation of PDC Guidelines, local PDC projects were expected to develop fully operational programs that would engage parents, teachers, and children in new patterns of interaction more conducive to the development of children's social competence. Possible new orders of parent-teacher-child interaction were implied, and in some instances actually described, in published documents, but no specific patterns--defined in terms of parent-child interaction, teacher-child interaction, or classroom process--were mandated. The few parent and teacher "outcomes" that were required by the Guidelines have already been considered in Chapter III, where they were treated as institutional features because they concerned parent and teacher roles within the institutional system: participation by parent and teacher representatives in program governance, internal assessment, and curriculum refinement; creation of classroom roles for parent volunteers and/or aides. Although ACYF's definition of the PDC model did not identify specific patterns of parent-teacher-child interaction necessary for the enhancement of children's social competence, the efficacy of the intervention was clearly dependent upon altering parent and teacher behavior toward children, for parents and teachers were the primary links, the social interface, between institution and child.

In this chapter, we address the third major research question guiding the evaluation (cf., Figure 1, Chapter 1), focusing on the domain of parent outcomes and leaving the teacher/classroom domain for the next chapter:

> Q2: *Did PDC parents exhibit attitudes, knowledge, and/or behaviors toward school and toward their own children that were different from those of non-PDC parents and, at the same time, compatible with the intention of the Guidelines?*

The findings of institutional impacts reported in Chapter III were not impressive taken as a whole: overall levels of Guideline implementation were typically moderate; at only two sites were PDC centers/schools clearly distinguishable from Comparison centers/schools with respect to the total configuration of prescribed institutional features. However, the Parent Involvement component of the model appeared to be significantly better implemented in PDC than Comparison centers/schools at six sites, suggesting that some PDC-favoring effects might be found in the parent outcome domain.

The remainder of this chapter is organized into three sections. The first describes measurement of parent outcomes; the second, our data analytic strategy; and the third, findings from analyses of parent outcome variables.

MEASUREMENT OF PARENT OUTCOME VARIABLES

ACYF placed very few requirements upon local PDC programs to demonstrate specific parent outcomes. Those areas for which there were specific expectations have already been considered in Chapter III: participation by parents of Head Start and elementary school children (a minimum of one each) on the PDC Council; participation by some parents as paid aides and volunteers in PDC classrooms. Other aspects of the mandated Parent Involvement component concerned affirmative action by centers and schools to involve parents more effectively in school life, and provision of training for parents that would help them to participate more effectively in the school setting and to more effectively support their children's learning at home.

In the absence of specific outcome objectives for individual parents, it was necessary to develop measures that would capture the potentially broad range of outcomes expected to result from local variation in implementation of the PDC Parent Involvement component. With guidance from the evaluation's Advisory Panel and ACYF representatives, the evaluators constructed a Parent Interview schedule to be used as the exclusive measure of individual parent and home environment variables. Although additional data on parent involvement in classrooms were collected in interviews with teachers and observations of classrooms, these data cannot be associated with individual parents or even be assumed to represent the behavior of parents of children in the evaluation cohort. Moreover, the variables derived from these data seemed most appropriately treated as indices of teachers' efforts to involve parents in classroom activities and teachers' approaches to managing parent involvement; thus, they have been grouped with teacher/classroom variables for primary analysis. In attempting to reach some conclusion about the impacts of PDC on parents and home environments, we shall, of course, try to integrate relevant findings from all domains.

Measures of Parent/Family Background

Very little information was obtained from or about parents at program entry. This situation resulted not by design or neglect but because of the extraordinary difficulty that we and many others experienced at the time in trying to secure approval of data collection forms by the U.S. Office of Management and Budget. In fact, it was not until the spring of 1979, when cohort children were completing first grade, that any significant number of parents of children in the evaluation cohort were interviewed for research purposes.

The only entry level data on parents/families available to the evaluation was information routinely recorded by local Head Start program staff when parents enrolled children at their center, as well as information that could be obtained by observing or talking with the child. Using these sources the following variables were measured: ethnicity, number of siblings, sex of child,

years of education of mother, child's previous preschool experience, presence or indication of a handicapping condition, and child's dominant language. These variables provided two direct measures of parent/family characteristics at program entry--mother's educational level and number of children in the family (number of siblings + 1)--and two indirect measures--ethnicity of family and dominant language of family (both inferred from characteristics of the child).

Additional family demographic data were collected in 1979, 1980, and 1981 as part of the Parent Interview, including information about employment status, family income, father's education, family structure (single/two-parent family), and updates of information about number of siblings and mother's educational level. Given explicit efforts by some local projects to improve the educational level and employment situation of participating parents, it was possible that later measures of employment, income, and education reflected program impacts. In addition, there was considerable missing data for measures of income and father's education, and information regarding occupational status was of very uneven quality. Of later demographic measures, only those pertaining to family structure were judged both insusceptible to program impacts and potentially useful additions to our small set of baseline measures as indicators of family demographic characteristics. Data from the Parent Interview contained information about family structure (single versus two-parent structure) at three points in time--spring 1979, 1980, and 1981. In deciding how best to represent this information for purposes of prediction and covariance analyses of parent and child outcomes, a series of exploratory regression analyses were performed for alternative constructions. By this process it was determined that available information about family structure was most parsimoniously represented by two dichotomous variables: families ever headed by a single parent versus families always headed by two parents (Family Structure 1); and families whose structure did not change over the course of three years versus families whose structure did change, either from one to two or from two to one parent (Family Structure 2).

In addition to direct questions about parent/family characteristics, the Parent Interview also asked parents whether they found it difficult to become involved in school life, and if so, why. Four response categories were judged to represent obstacles to involvement that were unlikely to have been directly affected by the intervention and that described aspects of the family's situation not reflected in other measures: parent must work; parent needs babysitter, or must take care of other children at home; family lives far from school; and family lacks transportation. Information from these four categories was used to construct a single variable for each interview year (1979-1981) representing "difficulty of involvement."

It is important to note that the parent/home environment variables considered in outcome analyses were not measured prior to the intervention, but only three years later as cohort children were completing first grade. Consequently, it was not possible (whether or not it would have been methodologically practical) to estimate change in parental attitudes, knowledge, and behavior over the full course of the intervention.

Outcome Data Collection and Availability

As indicated above all parent outcome variables were derived from Parent Interview data collected in the spring of 1979, 1980, and 1981. Interviewer training and general data collection procedures are described in Appendix E. The complete Parent Interview schedule is reproduced in earlier reports (e.g., Morris, Clement, & Rosario, 1980); only questions relevant to variables analyzed in the longitudinal evaluation are reproduced here (Appendix F).

Interviews were conducted only with parents of children in the evaluation cohort. For purposes of the longitudinal evaluation, only data for parents of children in the longitudinal analytic sample (n=326) were considered, since it was only these parents whose behavior might link institutional impacts to measured child outcomes. The number of parents for whom data were obtained at each point of measurement is reported for each parent outcome variable in Table IV-1. As can be seen, the actual number of parents for whom data were available was consistently smaller than the total number of children in the longitudinal analytic sample. Even when seven pairs of twins are taken into account (reducing the total potential sample to 319 parents or sets of parents/guardians), missing data, or measurement attrition, for the parent sample ranges from 14 to 19 percent except for baseline variables measured at entry into Head Start.

Measurement Reliability

The reliability of individual scores derived from interview data is seldom discussed in or addressed by educational research. Rather, one accepts the uncertain reliability of such measures and goes on.

Although the interview development effort in this evaluation was not explicitly concerned with estimating measurement reliability, the data collection design of the study allows us to examine the stability of parental responses to the same interview questions over three occasions--spring 1979, 1980, and 1981. Unlike the presumably stable individual traits measured by

Table IV-1

Available Data for Parent/Home Environment Variables
Considered in the Longitudinal Evaluation
Total Analytic Sample (N = 319)[1]

	BASELINE	LATER BACKGROUND & OUTCOMES		
	Head Start Fall 1976 N	G1 Spring 1979 N	G2 Spring 1980 N	G3 Spring 1981 N
Involvement in School/Program				
Frequency of Involvement	-	260	272	272
Breadth of Involvement	-	260	272	272
Parent as Educator				
Program-Induced Knowledge/Skill	-	260	272	272
Stimulation of Reading	-	260	273	272
Support for Learning	-	260	272	272
Communication about School	-	257	271	273
Background Variables				
Baseline Demographics	319	-	-	-
Family Structure	-	261	273	273
Difficulty of Involvement	-	261	273	273

[1]The longitudinal analytic sample contains 326 children, including seven pairs of twins. Thus, there are only 319 parents or sets of parents/guardians.

many psychological tests, the "true behaviors" that the Parent Interview was intended to measure might have changed dramatically from year-to-year due to variations in implementation of the PDC program, to change in a parent's employment status, to the birth of a new child, and so on. Thus, one would not expect to find very high correlations from one year to the next. However, one would expect to find some correlation, barring extraordinary changes in the environment. Interview-reinterview correlations are presented below as each variable is described, providing the only information we have about interview score reliability/stability.

Parent Outcome Variables

As indicated in Table III-1, outcome variables have been organized into two groups: *Involvement in School or Program* and *Parent as Educator*. These categories have a conceptual, rather than an empirical, origin, reflecting ACYF's implied objectives for parental behavior:

- that parents should become more effectively involved in all facets of school life; and
- that parents should become more effective in providing out-of-school support for their children's learning.

Descriptions of the outcome variables in each domain follow. Descriptive statistics for the items and the composite variables formed from them, item intercorrelations and item-composite correlations, and composite scale intercorrelations across time-points of measurement for each outcome variable scale formed are presented in Appendix F.

Parent Involvement in School or Program

Six questions of the Parent Interview asked parents to describe the *frequency* and *breadth* of their participation in school or PDC program affairs during each school year in which they were interviewed. A large number of categories were provided for classifying parents' responses to these questions and subquestions. Indeed, in an earlier report (Morris et al., 1980), some 50 discrete variables were derived from responses to these questions and analyzed separately, producing a complex array of findings that defied summarization. For this final, synthetic stage in the evaluation, steps were taken to reduce the number of discrete measures of frequency and breadth of parental involvement, by creating composite variables. Ultimately, the original 50 variables were represented by two composite scales--Frequency of Involvement and Breadth of Involvement. Initial derivation of these composites was logical rather than empirical; however, the empirical properties of items were closely examined before final decisions were made, to ensure that trends at the level of composite scales did not mask contrary trends at the item level.

Frequency of Involvement. The frequency of parental interaction with school was estimated using information from Parent Interview questions 3, 4, 5, and 8 (reproduced in Appendix F). Items 4, 5, and 8 were transformed into three-point scales classifying frequency as "never," "low," or "high"; item 3 was left as a two-point scale:

Item 3: Visited school in the past year? (originally coded yes/no)
1 = never visited
2 = visited at least once

Item 4: Number of times observed in child's class? (originally coded 0-99)
1 = never observed
2 = observed 1-3 times
3 = observed 4 or more times (high)

Item 5: Frequency of meeting or social activity attendance? (originally coded 0-5)
1 = never attended
2 = attended every few months or less (low)
3 = attended monthly or more often (high)

Item 8: Frequency with which parent worked (volunteer/pay) in child's school? (originally coded 0-5)
1 = never
2 = worked monthly or less (low)
3 = worked 2-3 times per month or more (high)

The definition of cutpoints for low versus high frequency was guided by examination of response distributions for the total sample. The composite variable--Frequency of Involvement--was formed using a set of logical rules, rather than simple addition, to combine the 4 component variables described above. These rules are detailed in Appendix F. The resulting composite variable had values ranging from 1 (no involvement) to 8 (very frequent involvement). In part because parents interacted with school in characteristically different ways, item intercorrelations were generally low (< .30); however, the stability of overall frequency of involvement from year to year was moderate (.48 for adjacent years; .33 from 1979 to 1981).

Breadth of Involvement. Breadth of Involvement was measured using information from Parent Interview questions 5 through 8 (Appendix F), describing the many different facets of parents' interactions with school. Observing in the classroom (question 4) was not included as a distinct facet, since all parents who reported observing in class also reported numerous other types of interaction with school. For example, in 1981 fully 98% of parents who observed also reported that they had spoken with the teacher about what their child "learns in school (Item 6a)". A content analysis of all types of involvement mentioned suggested that four fairly distinct facets were represented, each by multiple response categories associated with particular interview questions. Further analysis of the item sets associated with each major facet suggested that each set might

reasonably be reduced to four variables, some of which would be composites of several items. All items were coded 1/0 (mentioned by the parent or not); composite variables were coded 1 if any component had a value of 1. The 16 variables resulting from this process are described below by major interactional facet:

Item 5a: Type of Group Participation (originally 5 categories)

i. PTA/PTO/PAC meetings
ii. Training workshops or courses
iii. Council, committee, or task force meeting
iv. Social gathering or other unclassified activity

Item 6a: Topics of Conversation with Child's Teacher (originally 10 categories)

i. Educational issues (what child is learning, materials used, or parent's ideas about child's program)
ii. Behavioral issues (child's behavior in school, way teacher runs classroom, or classroom discipline)
iii. Working in classroom
iv. Other (general school activities, unspecified problems, or unclassified responses)

Item 7a: Roles of School Staff with Whom Parent Met (originally 7 categories)

i. Principal
ii. Health care staff or counselor/social worker
iii. Teacher other than child's class teacher
iv. Other (parent coordinator, PDC staff, or unclassified)

Item 8a: Work Functions Performed in School as Volunteer or Paid Aide (originally 8 categories)

i. Assisted teacher by working with children
ii. Assisted teacher by making materials, cleaning up, or helping with field trips
iii. Worked in playground, cafeteria, office, clinic, or library
iv. Worked on committees

The final composite variable--Breadth of Involvement--was constructed by summing values for the 16 variables described above, which produced a scale with values ranging from 0 to a theoretical maximum of 16. (Prior to analysis, 1 was added to each score for the composite producing a scale ranging from 1 to 17.)

Component variable intercorrelations were generally moderate to high among components representing the same facet, but low among components representing different facets. The stability of composite variable scores across time was quite high, ranging from .53 to .56 over the three year period for which data were available.

Parent as Educator

This outcome domain comprises four composite variables measuring parental knowledge and skills related to supporting children's learning and educationally relevant parental behavior toward (or interaction with) their children at home.

Program-Induced Knowledge or Skill. Two Parent Interview questions (10, a and b; reproduced in Appendix F) asked parents to evaluate the impacts of school or program activities upon their knowledge and skills related to their child's learning:

- Has school helped you learn how to help child with his/her school work? and
- Has school helped you know more about what child is learning in school?

Parent responses were coded in one of two categories: 1 = not helpful; 2 = helpful. The composite variable--Program-Induced Knowledge/Skill--was constructed by summing component scores, then subtracting 1 from the total. The composite variable scale ranged from 1 (not helpful in either case) to 3 (helpful in both cases).

Component variable intercorrelations were low, ranging from .28 to .34. The stability of composite scores from year to year was also low, ranging from .18 to .29. Scores were as likely to go down as up over time. Whether low stability reflects unreliability of parental responses, or real changes in parental opinion about the helpfulness of school/program over time, cannot be determined.

Stimulation of Reading. Because reading achievement had become a primary measure of children's success or failure in school, a special effort was made to obtain information about the extent to which children's home environments actively supported interest in and mastery of reading. The Stimulation of Reading variable was constructed from responses to 3 Parent Interview questions (12-14; reproduced in Appendix F):

- Availability in the home of books or magazines other than those the child brought home from school (coded yes/no);
- Frequency with which someone had read with the child at home during the past month (coded 1-5 on a relative frequency scale); and
- Who initiated reading with the child--no one, child or adult, both.

The third of these variables refers to parent-child interaction and, for that reason, confounds child and parent outcomes. However, the question asked about the role of the adult in stimulating the child's interest in reading and can appropriately be seen as a parent outcome if development

of interest in reading is viewed as a process in which interaction is essential. Stated differently, one would not expect a child to attempt to get an adult to read with him/her unless the child had previous enjoyable experiences reading with that adult. Thus, child initiation is likely to be an outcome of parent initiation, and a positive coding for the third variable legitimately reflects an outcome of parent efforts.

Responses to the three component questions were first ordered so that they had similar directions, then standardized. The final scale was formed as the mean of the three standardized values. Item intercorrelations generally increased over time, ranging from .08 to .52; composite scale correlations for successive years were .35 and .46.

Support for Learning. This scale provides an index of the frequency with which parent and child engaged in educational activities related to the acquisition of numeracy, oracy and literacy skills. The scale was constructed from responses to three Parent Interview questions (18 d, h and g; reproduced in Appendix F):

- How often does the parent play counting or word games with the child? (coded 1-4, relative frequency)
- How often does the parent work with the child on school-type activities? (coded 1-4, relative frequency)
- How often does the parent help the child with homework? (coded 1-4 relative frequency)

The latter two questions are obviously related and responses were highly correlated (ranging between .52 and .74 in different years) for parents who reported that their children brought homework home (88% of the sample by third grade, in 1981). In order to avoid weighing this activity twice in the scale, the information from these two items was combined into a single four-point scale by taking the more frequent (higher) of the two responses. The composite variable--Support for Learning--was formed as the sum of the composite score for the latter two questions and the score for the first question; resulting scores, after subtracting 1 from all values, ranged from 1 to 7. Item correlations, excluding those between the second and third questions, ranged between .28 and .42; composite scale correlations across successive years were .35 and .38.

Communication about School. This is a composite variable constructed by averaging scores for responses to two Parent Interview questions (18 e and f; reproduced in Appendix F):

- Frequency with which parent talks with child about what goes on in school (coded 1 to 4, relative frequency);
- Frequency with which the parent talks with child about child's feelings toward school (coded 1 to 4, relative frequency).

Unlike other outcome variables in this domain, Communication about School does not directly measure parental support or capacity to support the child's learning. Rather, this composite variable indexes the level of interest in and value placed upon schooling in the home environment as expressed in conversations between parent and child. However, one way of supporting school-related learning is to nurture the child's interest in schooling by legitimating its importance within the culture of home and community. Component variable intercorrelations at the three data collection points ranged from .42 to .49; the stability of composite scores over the three periods ranged from .28 to .36.

Overview of Parent Outcome Measures

There is no question but that measurement of parent outcomes was less thorough than we would have liked. Nevertheless, if parents' self-reports can "be believed," the available data should reflect any major impacts that may have occurred in areas of particular relevance to PDC. The question of whether parents' self-reports are believable--i.e., reliable and valid--cannot be answered definitively. Evidence of generally moderate (>.30) and occasionally strong (>.50) correlations between responses from one annual interview to the next indicates that parental responses were definitely not "random." Considering the many uncontrolled factors in family life that might have altered actual levels of involvement or the character of parent-child interactions, the stability of parental responses is really quite impressive. Whether the information obtained was not only relatively consistent across time but valid cannot be known since independent measures of parent involvement and parent-child interaction are not available. But whatever the validity of data obtained from particular parents, there is no evidence to suggest that parental responses were systematically biased in such a way as to prejudice estimates of program effect based on comparisons of group means.

ANALYTIC DESIGN

The basic methods used to analyze parent outcomes have already been described in detail in the methods section of Chapter II--The Evaluation Design. Findings from Designs 1 and 2 are presented in Appendix I; the synthesis of findings from the two designs is presented in this chapter. The only feature of the analytic design for parent outcomes that requires explication here is the specific set of predictor variables used as covariates.

Covariates Incorporated in the Analytic Design

As indicated in a previous section of this chapter (Measures of Parent/Family Background), relatively little information could be obtained about the pretreatment characteristics of families whose children were involved in the evaluation. The most significant limitation on background information was the unavailability of specific data on socioeconomic status: other than knowing that all PDC and Comparison group families had to have been relatively poor for their children to have qualified for Head Start, we were in the dark. Need for reasons already mentioned SES data collected in 1979, 1980, and 1981 could not be used to identify much less control for potential group nonequivalence.

The variables finally selected for use as covariates in Designs 1 and 2 represent the best set of predictors we could assemble under the circumstances. They include both direct measures of family characteristics and indirect estimates of family characteristics based on measures of children. Each of the nine variables incorporated in the covariate set is listed below and described as necessary:

- Mother's educational level (0-13+ years of schooling).
- Family structure I (two-parent structure throughout the evaluation/one-parent structure at some point in the evaluation).
- Family structure II (stable one- or two-parent structure/ unstable structure). These two dimensions of family structure appeared to capture the important, predictive variance in family structure over the last three years of the project for which information was available.
- Family ethnicity (Anglo/Other). Family ethnicity was inferred from the ethnicity of the child. Ethnic status was represented as a dichotomous variable contrasting "Anglos" with everyone else because this distinction (rather than the five ethnic categories into which data were originally coded) was most consistently predictive of later outcomes.
- Dominant language (Spanish/English). The family's dominant language was inferred from information about the child.

- Child's prior preschool experience (None/Some). Since a child's previous preschool experience was determined by parental decisions, whether or not a child had attended preschool prior to Head Start was thought to say something about the family. Just what it says is not clear. This information was collected by Head Start staff when children were enrolled in 1976 and it appears to confound prior "day care" with prior "preschool" experience. "Prior preschool experience," whatever its meaning for the child or family, correlates negatively with later experience.

- Sex of child. Since parents frequently interact differently with and act differently on behalf of children of different sexes, this variable was included as a covariate both to reduce outcome variance (in order to increase the statistical power of our tests) and to control for any bias in outcomes arising from differences in the sex ratios of small site-level child samples.

- Impediments to parent involvement in school affairs. In each of the three years (1979-1981) that interviews were conducted parents were asked whether there were any specific obstacles to their involvement in school affairs. Covariate scores were derived from principal component factor analyses of four response categories: must work; needs babysitter or must care for other children at home; lives far from school; lacks transportation. Factor scores were estimated for each year's interview data, and only scores obtained in the same year as the outcome being analyzed were incorporated in the covariance designs.

- Child's test performance during the Head Start year--2 factor scores. The derivation of two factor scores from 9 baseline measures repeated in fall and spring of the Head Start year is described in Appendix H. These two scores appear to capture most of the important, predictive variance in baseline measures. They are included as covariates in analyses of parent outcomes because they were thought to be influenced by and therefore indicators of the educational/developmental support provided by parents during their children's early years (something not directly measured in the evaluation).

Together these covariates fairly consistently explained from 20 to 30% of the variance in parent outcomes in multiple linear regressions performed in Design 1 and Design 2 analyses (see tables summarizing effects for Designs 1 and 2 in Appendix I).

FINDINGS

Again, the reader is directed to Chapter II--Data Analysis Methods Used in the Final Evaluation--for specific guidance in interpreting the presentation of findings in this and remaining chapters.

Net effects of consequence are summarized graphically in Figure IV-1 for each outcome domain: Parent Involvement in School/Program and Parent as Educator.[1] Sites where consequential net effects were found are represented by numerals placed above the horizontal bar indicating the range of site-level net effects for each domain. Site position on the horizontal axis indicates the magnitudes and directions of observed net effects. Negative net effects (those appearing to the left of the zero-effect midline) are Comparison-favoring; positive net effects (to the right of the midline) are PDC-favoring. Net effects from comparisons of aggregate PDC and Comparison samples are represented by the letter "A" (appearing below the horizontal bar) whether or not the effect was judged consequential.

All estimates of net observed and net predicted effects are summarized in Table IV-2; net effects judged consequential and, thus, reported in Figure IV-1 are highlighted. Site-level and aggregate findings from the synthesis of Designs 1 and 2 are presented in Tables IV-3 through IV-5 for *probable* effects, *probable and possible (?)* effects, and *probable, possible (?), and less possible (??)* effects, respectively. These tables show (at three levels of probability) the patterns of findings across time and across variables within outcome domains upon which estimates of net effect were based for each site and the total sample. Primary judgements of the consequentiality of findings were based on the patterns shown in Table IV-4 for *probable and possible* effects. Detailed results of all Design 1 and 2 analyses are reported in Appendix I along with observed, predicted, and ANCOVA-adjusted means for all outcome variables by group by site and by group aggregated across site.

Findings are first summarized briefly, then discussed in some detail. An effort is made to explain findings for parents by reference to institutional findings reported in Chapter III. The fit between parent outcomes and outcomes for teachers/classrooms and children will be considered in subsequent chapters.

---

[1] As explained in Chapter II, net effects were computed by subtracting the number of negative (-) findings from the number of positive (+) findings in a domain, then dividing by the total number of findings (+, -, and null) in the domain. Net effects of .25 or greater, based on pooled *probable* and *possible (?)* findings, were considered "consequential". Reported net effects include *less possible (??)* findings only when the addition of these findings increased the directional strength of the net effect based on *probable* and *possible* findings.

Summary of Parent Outcome Findings

The evaluation of parent outcomes was guided by the following question:

> *Did PDC parents exhibit attitudes, knowledge, and/or behaviors toward school and toward their own children that were different from those of non-PDC parents and, at the same time, compatible with the intentions of the Guidelines?*

The answer: at several sites, *yes;* in the aggregate, *no.*

Site-Level Effects

Parent Involvement: The PDC intervention appears (Figure IV-1) to have increased parent involvement above levels prevailing in local non-PDC schools at two sites--Sites 1 and 10. At one site--Site 6-- Comparison schools seem to have been more successful than PDC schools in their efforts to involve parents in school affairs. These and other findings reported here are, of course, restricted to parents of children in the longitudinal analytic sample.

Parent as Educator: PDC programs at two sites--Sites 1 and 7--appear to have increased parents' knowledge, skills, and/or action in support of their children's school-related learning. At one site--Site 3--findings in this domain favored Comparison parents and, by inference, Comparison schools.

Aggregate Effects

There was no indication of consequential effects in either domain when data for the aggregate PDC and Comparison samples were analyzed. This finding is certainly compatible with the mixed directions of net effects at the site level (Table IV-2).

Discussion of Findings

The small number of consequential net effects found in site-level analyses does not appear to have resulted from the occurrence of contradictory (+/-) findings across variables or across time within domains, but from the occurrence of relatively few *probable* and *possible* effects in either direction at most sites (Table IV-4). Although low statistical power may have kept us from identifying consequential effects at some sites, it would not appear to have obscured any important directional trend across sites given the absence of consequential net effects for the aggregate PDC and Comparison samples. Furthermore, there is no indication (Tables IV-3 through IV-5) that our data analytic methods obscured any important temporal trend--i.e., it does not appear that PDC/Comparison differences (positive or negative) were more likely to occur in 1979 (kindergarten year) than 1981 (third grade year), or *vice versa.*

Figure IV-1

Parent Outcomes
Consequential Net Effects by Domain and Site

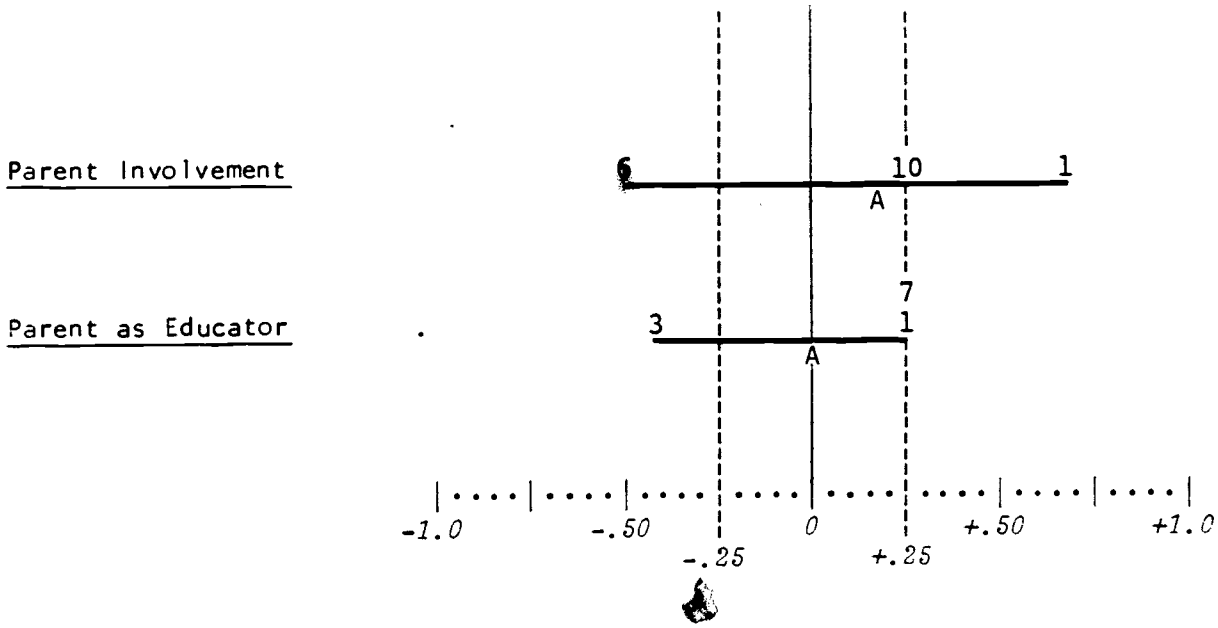


Table IV-2

Parent Outcomes
Summary of Net Observed Effects and Predictions by Site
and for Aggregate PDC and Comparison Samples at Three Levels of Probability

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Agg
Involvement	Probable	.00	.00	.00	.00	.00	-.17	.17	.00	.00	.00	.17
	Prob/?	.33	.00	-.17	-.17	.00	-.33	.17	-.17	.17	.25	.17
	P/?/??	.67	.00	-.50	-1.0	.00	-.50	.33	-.17	.17	.25	-.17
	Predicted	-.33	.00	.33	1.0	.00	.67	-.17	.00	.00	.00	.67
Parent as Educator	Probable	.08	-.08	.00	.00	.00	-.08	.00	.08	.00	.00	.00
	Prob/?	.25	-.17	-.25	.00	.00	-.17	.25	.17	.08	.12	.00
	P/?/??	.25	-.17	-.42	.08	.00	-.25	.25	.17	.08	.12	.00
	Predicted	.00	.00	-.08	.00	.00	-.08	-.17	-.08	.00	.00	.00

Table IV-3
Summary of Probable Effects for Parent Outcomes[1]

Outcome Domain and Measures	Sites										
	1	2	3	4	5	6	7	8	9	10	Agg
	123	123	123	123	123	123	123	123	123	12*	123
Involvement in School											
Frequency of Involvement	000	000	000	000	000	-00	+00	000	000	00	000
Breadth of Involvement	000	000	000	000	000	000	000	000	000	00	+00
Parent as Educator											
Program-induced Knowledge/Skills	000	000	000	0-0	000	000	000	+00	000	00	000
Stimulation of Reading	000	000	000	000	000	00-	000	000	000	00	000
Support for Learning	000	000	000	00+	000	000	000	000	000	00	000
Communication About School	00+	0-0	000	000	000	000	000	000	000	00	000

Table IV-4
Summary of Probable and Possible (?) Effects for Parent Outcomes[1]

Outcome Domain and Measures	Sites										
	1	2	3	4	5	6	7	8	9	10	Agg
	123	123	123	123	123	123	123	123	123	12*	123
Involvement in School											
Frequency of Involvement	000	000	000	0-0	000	-0-	+00	00-	000	+0	000
Breadth of Involvement	++0	000	-00	000	000	000	000	000	+00	00	+00
Parent as Educator											
Program-induced Knowledge/Skills	000	000	00-	0-0	000	000	000	+00	000	00	000
Stimulation of Reading	0+0	000	000	000	000	00-	00+	+00	000	0+	0+0
Support for Learning	000	000	000	00+	000	0-0	0+0	000	000	00	000
Communication About School	0++	0--	--0	000	000	000	0+0	000	0+0	00	00-

Table IV-5
Summary of Probable, Possible (?), and Less Possible (??) Effects For Parent Outcomes[1]

Outcome Domain and Measures	Sites										
	1	2	3	4	5	6	7	8	9	10	Agg
	123	123	123	123	123	123	123	123	123	12*	123
Involvement in School											
Frequency of Involvement	0++	000	0-0	---	000	-0-	+00	00-	000	+0	0--
Breadth of Involvement	++0	000	--0	---	000	-00	+00	000	+00	00	+00
Parent as Educator											
Program-induced Knowledge/Skills	000	000	00-	0-0	000	000	-00	+00	000	00	000
Stimulation of Reading	0+0	000	0-0	+00	000	00-	00+	+00	000	0+	0+0
Support for Learning	000	000	00-	00+	000	0--	0+0	000	000	00	000
Communication About School	0++	0--	--0	000	000	000	++0	000	0+0	00	00-

[1]Derived from the synthesis of Designs 1 and 2. Reported for each site and the aggregate sample (Agg) by year (1=grade 1/1979; 2=grade 2/1980; 3=grade 3/1981).
*Not enough parent responses were available from Site 10 in 1981 for analysis.

For readers familiar with Interim Report X (Morris, Clement, & Rosario, 1980; Wacker, Clement, Morris, & Rosario, 1980) it should be pointed out that the findings reported here essentially confirm the findings reported through the first grade year. This is true in spite of the fact that the longitudinal analytic sample is considerably smaller (by 29%) than the first grade sample, and in spite of substantial differences in the construction of analytic variables. Confirmation can be found in the aggregate test of Breadth of Parent Involvement at first grade (Table IV-4) and in aggregate tests of teacher attitude toward parent involvement and actual involvement of parents in the classroom (Chapter V, next).

Although it might seem that differences in Parent Involvement would be associated with (perhaps even responsible for) differences in Parent as Educator outcomes, this was not generally the case. Of the five sites where consequential net effects were found in at least one domain, at only one--Site 1--were consequential effects found in both domains. Since Site 1 net effects were in the same direction (positive), they do not contradict the expectation that outcomes for the two domains would be related; however, the pattern across sites suggests that the relationship between outcome domains was at best weak. Correlations of variables in the Parent Involvement domain with variables in the Parent as Educator domain (and of factor scores computed for the variables in each domain) were less than .25 in every instance (based on analyses of data for the total sample), further indicating that mere involvement of parents (as measured) does not lead inevitably to improved parental knowledge, skill, or action in support of children's school-related learning.

In Chapter III--PDC's Influence on Institutions--we concluded that implementation of the Parent Involvement component of the PDC Guidelines had produced appreciable differences between PDC and Comparison schools/centers at six sites--Sites 1, 2, 5, 6, 7, and 9. These differences involved presence/absence or degree of implementation of "mechanisms" (staffing, training, decision-making groups) that might be expected to influence the intensity and scope of parent involvement in school/program affairs. All differences favored PDC. Based on these findings we speculated that PDC-favoring differences in parent outcomes might also be expected for those sites. The relationships between expectations and findings are illustrated below:

Observed Differences (Net Effects) in Parent Outcomes	Expected Differences in Parent Outcomes: +	Expected Differences in Parent Outcomes: 0
+/+	1	
+/0	7	10
0/0	2,5,9	4,8
-/0	6	3
-/-		

The numeric entries represent sites. "+/+" indicates that consequential net positive effects were found for both parent outcome domains; "+/0", that a net positive effect of consequence was found for one domain, while no consequential effect was found for the other; and so on.

The implications of the table are clear: the findings for parent outcomes did not fulfill our expectations. Though the fit between institutional and parent findings at Site 1, and to a lesser extent at Site 7, leads us to feel more confident that the PDC-favoring parent outcome findings at these sites represent true program effects, the overall lack of fit between institutional and parent outcomes raises serious questions about the adequacy of the intervention model and our evaluation of it. We may have inadequately measured implementation of PDC Guidelines, or relevant parent outcomes, or both. We may have failed to identify real program impacts when they occurred and/or to reject findings that were the products of pretreatment factors. The causal linkages implied by the intervention model may not exist in the world, or perhaps the specification of linkages was insufficient to bridge the gap between prescribed institutional features and parental behavior.

Yet another possible explanation of the lack of fit between institutional and parent outcomes is that the tiny sample of PDC and Comparison parents included in the longitudinal evaluation was not representative of the universe of parents of children attending PDC and Comparison schools. After-all, parents included in the evaluation were at most 5% (and often a much smaller proportion) of parents whose children attended designated PDC and Comparison schools, and typically less than 25% (often less than 5%) of parents whose children were enrolled in the same classrooms. In short, the analytic sample may be an odd lot of parents, whose involvement in school and educational initiatives in the home were not typical of parents in the universe served by PDC and/or Comparison schools. (As mentioned earlier, we even have doubts that parents and children in the longitudinal analytic sample adequately represent the sample initially selected for inclusion in the evaluation because of extremely high and possibly nonrandom attrition over the course of the project.) The possibility that PDC had positive impacts on parents in general, whether or not it affected parents in the analytic sample, will be considered further in the next chapter where we examine teachers' efforts to promote parent involvement and associated levels of involvement in the classroom process by parents "in general."

V

TEACHER/CLASSROOM OUTCOMES

In this chapter we address the third research question posed in Chapter II (cf., Figure II-1):

> Q3: Did PDC teachers exhibit attitudes, knowledge, and/or behaviors toward children that were different from those of non-PDC teachers and, at the same time, compatible with the intentions of the Guidelines?

In addressing this question we consider not only direct measures of teacher attitudes, knowledge, and behavior, but also measures of what might more appropriately be described as classroom process/structure and content--variables that may be very strongly influenced by teachers, but that are actually the products of interactions among all members of the class group and between this group and the physical environment of the classroom.

Although ACYF's definition of the PDC model did not prescribe specific patterns of teacher behavior toward children or specific classroom processes, the model assumed that positive impacts on children would be achieved by way of changes in teacher behavior and classroom process, together with changes in parents' behavior, along lines implied by the Guidelines and by statements of PDC's broad goals. The intervention model further assumed that implementation of PDC Guidelines, defined in terms of institutional features, would set in motion forces sufficient to produce such changes in teacher/classroom variables. Although overall findings of Guideline implementation and resulting differences between PDC and Comparison institutions (cf., Tables III-5 and 6 in Chapter III) would not lead one to expect many or strong program effects in teacher/classroom outcomes, findings of PDC-favoring differences in the implementation of the Parent Involvement component of the Guidelines at six sites suggested that some effects might be found in outcomes representing teachers' efforts to promote parent involvement in the classroom. Moreover, whether or not PDC Guidelines were consistently well implemented, the PDC projects at particular sites may have provided local innovators with opportunities to develop new educational approaches that would be revealed by analyses of teacher/classroom outcomes. Thus, we approached this outcome area with few strong expectations but open minds.

The first section of this chapter describes measurement of teacher/classroom outcomes; the second, unique features of the data analytic design; and the third, findings from the analysis.

MEASUREMENT OF TEACHER AND CLASSROOM ENVIRONMENT VARIABLES

ACYF placed few requirements on local PDC programs to demonstrate specific teacher outcomes. Those areas in which there were specific requirements have already been evaluated in Chapter II: participation by teachers (at least one Head Start and one elementary school teacher) on the PDC Council; participation by at least some teachers in formal processes of internal program assessment and curriculum refinement; acceptance (in some unspecified measure) of parents in the classroom in roles as volunteers or aides. In addition to these required outcomes, several expectations of teachers were strongly implied by various required program components as well as in ACYF's public discussion of PDC goals and possible methods. Specifically, it was clear that ACYF expected PDC teachers to become actively involved in reducing discontinuities between home and school, by seeking and encouraging extensive and substantive parent participation in the classroom. Moreover, teachers were clearly expected to take real initiative to individualize instruction for all students and to integrate multicultural education, health and nutrition education, and community resources into their instructional programs. Although ACYF also expected local PDC programs to develop more effective approaches to basic skills instruction, no particular approach was specified.

With guidance from the evaluation's Advisory Panel and ACYF representatives, the evaluator attempted to develop instruments that would measure a broad range of teacher/classroom outcomes, including not only those areas in which fairly specific outcomes were expected, but also areas such as class management and allocation of learning time, where local programs might or might not have developed distinctive approaches within the mandated PDC institutional framework. Two instruments emerged from this development effort--the PDC Teacher Interview and PDC Classroom Observation System (COS).

No information was obtained about the attitudes or behavior of teachers prior to or at entry to the program. All teacher/classroom data were collected in spring 1979, 1980, and 1981, when (the bulk of) evaluation cohort children were in first, second, and third grades, respectively. Included in the Teacher Interview were selected biographical questions and observations by the interviewer of teacher characteristics (gender and ethnicity). The biographical data collected included the following: total years of prior teaching experience; total years of prior teaching experience at current grade level; number of years teaching at current school; highest degree attained; how came to teach in current school/program (chance/own choice/invitation).

Outcome Data Collection and Availability

All teacher/classroom outcome variables were derived from data collected with two instruments--Teacher Interview and Classroom Observation System (COS)--administered each spring from 1979 through 1981. Interviewer and observer training, and general data collection and processing procedures are described in Appendix E. The complete Teacher Interview schedule has been reproduced in earlier reports (e.g., Wacker, Clement, Morris, & Rosario, 1980), while the various observation schedules and ratings scales constituting the COS are presented in the PDC Classroom Observation Manual (Wacker, Rosario, Diamondstone, & Smith, 1981). Only interview questions and particular rating scales that were actually used to generate outcome variables, and the COS observation schedule are reproduced here (Appendix G).

Observations were made only of classrooms attended by children in the evaluation cohort; interviews were conducted with all teachers in the evaluation cohort plus one randomly selected teacher from each grade level not already represented in the cohort sample in each PDC school/center and matched Comparison school/center. Unlike parent and child samples, the teacher sample comprised largely different teachers each year as the evaluation cohort moved through the early elementary grades. A small number of teachers were observed in more than one year, either because they changed grade levels in sync with the cohort or because a cohort child in their class did not: 45 teachers were observed in both 1979 and 1980; 31 in 1980 and 1981; and 10 in all three years. A larger number of teachers were interviewed in more than one year because interview data were collected from teachers, chosen by random selection from a very small universe, who were not teaching cohort children: 81 teachers appeared in both the 1979 and 1980 interview samples; 69 in 1980 and 1981; and 52 in all three years. For two reasons it was decided to restrict teacher/classroom outcome analyses to data collected only from teachers of cohort children--i.e., teachers who were both interviewed and observed. First, with this restriction, analytic samples for outcome variables derived from the Teacher Interview and the COS would be the same, greatly facilitating the integration of teacher/classroom findings. Second, by restricting the sample to teachers of cohort children, teacher/classroom findings might be directly interpreted (various other things "being equal") as linking institutional with child impacts. In keeping with the latter rationale, the sample was further restricted to teachers of children in the longitudinal analytic sample.

Thus restricted, the maximum possible teacher/classroom sample equalled the number of classrooms in which children from the longitudinal analytic sample were enrolled each year from 1979 through 1981. The number of teachers/classrooms for whom/which data were actually obtained at each point of measurement are reported in Table V-1, which also lists each of the outcome variables organized into conceptual categories. As the table shows, measurement-related attrition for this sample is low: overall, across the three years, the percentage of data missing ranges from 1 to 6%, and for none of the annual measurements is it higher than 9%.

Table V-1

Available Data for Teacher/Classroom Outcome Variables
Considered in the Longitudinal Evaluation

	1979-1981 N(max)=392	1979 N(max)=129	1980 N(max)=135	1981 N(max)=128
Promotion of Parent Involvement				
Attitude toward Parent Involvement	380	125	131	124
Amount of Home Visiting	385	126	133	126
Parents as Teachers	370	126	127	117
Minutes of Parent Time in Class	385	124	135	126
Classroom Environment				
Stimulating & Attractive Physical Environment	384	124	135	125
Supportive & Enthusiastic Climate	383	123	135	125
Educational Management				
Maintenance of Orderly Classroom Process	384	123	135	126
Management of Information for Individualization	385	128	132	125
% of Time Spent Teaching Children	385	124	135	126
Level of Teacher/Child Involvement in Learning	383	123	135	125
% of Child Time Engaged in Any Educational Activity	385	124	135	126
% of Child Time Engaged in Learning with High Attention	385	124	135	126
PDC-Encouraged Instructional Content				
Multicultural Instruction	387	127	133	127
Health/Nutrition Instruction	387	127	133	127
Use of Community Resources	388	128	133	127
Learning Time				
Minutes/Child Available for Learning	385	124	135	126
Minutes/Child Allotted to Specific Academic Activity	385	124	135	126
Minutes/Child Allotted to Math	385	124	135	126
Minutes/Child Allotted to Reading/Language Arts	385	124	135	126
Minutes/Child: Math with High Attention	385	124	135	126
Minutes/Child: Reading/Language Arts with High Attention	385	124	135	126

Instrumentation

Teacher/classroom outcome variables were derived from responses to the PDC Teacher Interview and data generated by the various component instruments of the Classroom Observation System (COS). Before launching into a discussion of specific outcome variables, the presentation of which is not organized by instrument but by conceptual grouping, it seems advisable to describe the data collection instruments in some detail.

PDC Teacher Interview

For the most part, the Teacher Interview was straightforward, as the questions reproduced in Appendix G demonstrate. However, one section of the interview departed significantly from the structured approach followed in other sections. In that section, the interviewer engaged in a semistructured conversation with the teacher in an effort to glean insights into the teacher's instructional approach, particularly individualization of instruction, by analyzing the teacher's comments about actual students in her class. This conversation was guided by the following standard statements and questions:

> *Now I would like to find out a little about the approach that you use in your class to teach such things as language arts, number concepts, and so forth. Think of two average children in your class, one boy and one girl, who are not the best or the worst students. Let's call them _____ and _____.*
>
> *First, in the beginning of this year, how did you learn _____ and _____'s particular interests, needs and abilities in such areas as language arts and number concepts? Please be as specific as possible? Did you use any tests? Speak to others about them? Look for anything in your observations?*
>
> *Has your opinion about their needs or abilities changed since the beginning of the year? If so, on what information was your new opinion based?*
>
> *Could you tell me a little about what you are planning to do with your class tomorrow in the areas of language arts and mathematics? Specifically will _____ and _____ be doing similar or different things?*
>
> *Let's imagine that you had recently learned you were going to be away for the next month and I am the substitute who will be filling in for you. How would you describe your strategy for teaching language arts and mathematics?*
>
> *Supposing still that I were your substitute, what would you tell me specifically about _____ and _____? What records or other information could you share with me?*

immediately after finishing the interview, interviewers completed a set of five-point rating scales defined at their extremes by contrasting behavioral descriptions. Most of these scales were intended to measure some aspect of individualization of instruction, broadly construed. These ratings provided data for one of the outcome variables described later in this section--Management of Information for Individualization.

PDC Classroom Observation System (COS)

The Classroom Observation System (COS) was developed specifically for the PDC evaluation. It was designed to provide a comprehensive picture of the PDC and Comparison classrooms attended by children in the evaluation cohort. The COS comprised several different instruments:

- Classroom Activities Record (CAR) Observation System;
- Classroom Activities Record (CAR) Global Ratings;
- Classroom Environment Observation (CEO) Global Ratings;
- Focused Observation Classroom Climate (FOCC) Global Ratings;
- Focused Observation Periodic Ratings (FOPR).

Two complete days were required to collect COS data for a single classroom. CAR observations and ratings were made on the first day; FOCC ratings and FOPR, on the second; and CEO ratings, during spare moments on either or both days. Brief descriptions of each instrument follow.

CAR Observation System. The CAR observation system was designed to provide information about teacher management of instruction, the social organization of the classroom, and the learning experiences of children. Observations were recorded every five minutes for the entire period that the class was in session on the day selected for observation. Activities occurring outside the classroom--lunch, recess, gym, music, assemblies--were not observed. Observation days were selected in cooperation with class teachers to represent "normal" days. CAR observations measured the behavior of individual teachers but not individual children. Child behavior was characterized at the level of instructional groups, ranging from the whole class through any number of smaller groupings to unidentified individuals working independently. Coding forms (called "pages"; 1 page = 5 minutes of observation), coding category definitions, a description of data reduction procedures, and evidence of interobserver reliability are presented in Appendix G.

CAR Global Ratings. At the end of the CAR observation day, the observer completed a set of global ratings. For each scale observers were asked to select one of five points that best described the classroom in question. Extreme scale points (1 and 5) were defined by contrasting descriptive statements. The number of rating scales in the CAR set ranged from 46 in 1979 to 22 in 1981. Only 15 scales that were used consistently from 1979 through 1981 were considered for analysis in this evaluation; of these, only 8 were found to be sufficiently reliable (as estimated by inter-rater agreement) to merit analysis. These 8 scales and evidence of their inter-rater reliability are presented in Appendix G.

CEO Global Ratings. CEO ratings involved judgments about the physical environment of each classroom. Ratings were made on either or both days of observation when the class was not in session. The same 12 scales were rated from 1979 through 1981. All were five-point scales defined at their extreme points by contrasting descriptive statements. Only three scales were found to be sufficiently reliable and interpretable to be used in the evaluation. These scales and evidence of inter-rater reliability are presented in Appendix G.

Focused Observation Ratings. The second day of observation in each classroom was devoted to observations intended to measure the social-emotional climate of each classroom (FOCC) and to observations of specific aspects of teacher's efforts at intellectual stimulation and class management (FOPR). In 1979, climate, stimulation, and management variables were observed and rated separately during successive 15-minute periods that formed a 45-minute cycle repeated from the beginning to the end of the class day. At the end of the day, observers made global ratings of each climate, stimulation, and management variable. In 1980 and 1981, the observation cycle was redefined, requiring observers to devote 10 minutes to observation of climate variables (with note taking, but without rating) followed by 10 minutes observing stimulation and management variables, then 5 minutes rating stimulation and management variables. This 25-minute cycle was repeated throughout the class day. At the end of the day, global ratings were made of classroom climate but not of stimulation and management.

For this evaluation, end-of-day global ratings of classroom climate (FOCC) and average periodic ratings of stimulation and management (FOPR) from 1979 through 1981 data collections were considered. Only 4 of 25 climate (FOCC) ratings were found to be sufficiently reliable and interpretable for analysis. Eight of 11 management variables (FOPR), but none of the stimulation variables, met criteria for inter-rater reliability and interpretability. All 12 scales included in the classroom measurement battery are presented in Appendix G together with inter-rater reliability data.

Measurement Reliability

As in the case of outcome variables derived from the Parent Interview, we have sought to glimpse measurement reliability in the teacher/classroom domain by correlating outcome scores from one year to the next for the small number of teachers and classrooms appearing in our samples more than once. These correlations are presented below, along with variable descriptions, as potential evidence of the reliability of measures and the stability of the phenomena measured.

In addition to these rather tenuous estimates of stability, we present estimates of inter-observer reliability--percentage agreement and/or intra-class correlations--for all variables derived from the COS. These estimates, based upon a reliability study conducted during observer training in 1981, were also used to screen variables for inclusion in the longitudinal evaluation. The design of the reliability study and its results are summarized in Appendix G.

Teacher/Classroom Outcome Variables

As indicated in Table IV-1, outcome variables have been organized into five groups or domains:

1. Promotion of Parent Involvement
2. Classroom Environment
3. Educational Management
4. PDC-Encouraged Instructional Content
5. Learning Time

Domains 1 and 4 represent outcome areas for which ACYF had specific expectations. The remaining categories represent areas of potential program effect given stated PDC objectives pertaining to more effective basic skills instruction, and the implementation of Guideline-prescribed institutional/organizational mechanisms for the realization of these objectives. The variable groupings are conceptual rather than empirical; however, variable intercorrelations within domains were generally statistically significant and positive, sometimes not significant, but never significant and negative.

Promotion of Parent Involvement

The institutional features prescribed by the PDC Guidelines were clearly intended to improve teachers' attitudes toward parent participation in the formal educational process, to increase teachers' efforts to involve parents in the classroom, and ultimately to produce higher levels of parent involvement. The mechanisms for accomplishing attitudinal change and increasing teachers' knowledge of how to involve parents were formal training, participation in planning and decision-making groups with parents, specialist support for teachers and parents working together, and urging from project staff and peers. The five outcome variables included in this category were intended to measure teachers' attitudes toward, efforts to promote, and success in achieving parent involvement in their classrooms.

Attitude Toward Parent Involvement. The teacher's attitude toward parent involvement was measured by responses to a question in the Teacher Interview (question 17; reproduced in Appendix G) which asked teachers to list the advantages of involving parents in school. Parent responses were coded into ten categories. These were recategorized into two: one representing advantages or benefits to the teacher; the other, advantages or benefits to parents and children. Responses included in the former category represent a fairly traditional view of parents as extra hands in the service of teachers; the latter category comprises responses that acknowledge the special contribution parents can make to the educational process and/or the special benefits a child or parent may reap:

Parents as Extra Hands

- Another adult in the class helps with discipline and classroom management.
- Another adult allows the teacher to individualize instruction more effectively to meet the different needs of children.
- Another adult helps meet the special needs of handicapped children.
- Another adult gives the teacher more time to plan and observe.

Unique Contributions/Benefits

- Parents bring special skills with them that can be shared with children.
- By becoming familiar with school activities parents are able to do more for their children at home.
- Parents know the language and culture of the children and can bring that perspective to the school.
- Parent involvement creates a bridge between school and community. Involvement creates a better understanding among parents of school life, of the teacher's job; teachers get a better understanding of the parent's view.
- Parent involvement increases the child's self esteem. The child feels important when the parent comes to school; it makes the child more interested in school.
- Parent involvement increases the parent's understanding of the child's problems: how the child relates to peers; his/her work in school.

A composite variable--Attitude toward Parent Involvement--was constructed according to the following rules: 0 = no advantages mentioned; 1 = only advantages in category 1 (Parent as Extra Hands) mentioned; 2 = advantages in category 2 (Unique Contributions/Benefits) mentioned; 3 = advantages of both kinds mentioned. Since most teachers mentioned relatively few categories, item intercorrelations were not computed. Descriptive statistics and year to year correlations of composite scores are reported in Appendix G. Stability of the composite variable was low to moderate: correlations over time were .28 (1979-1980) and .32 (1980-1981).

Amount of Home Visiting. The number of children's homes visited by the teacher during the school year was considered indicative of initiative taken to bridge the gulf between home and school and to involve parents more effectively in the formal educational process. The number of homes visited was measured by a question in the Teacher Interview (question 10; reproduced in Appendix G). The responses ranged from none to a maximum of 30. The variable was rescaled into four ordered categories: no home visits, one to five home visits, six to fifteen home visits and sixteen or more home visits.

Descriptive statistics and year-to-year correlations for this variable are reported in Appendix G. Stability was high: correlations between years ranged from .65 to .73.

Parent as Teacher. Teachers who indicated that at least one parent had visited their classroom (Teacher Interview question 13) were asked what proportion of parent visitors "helped by working with children" (question 14-1; both questions are reproduced in Appendix G). Three response options were offered: 1 = none or few; 2 = some; 3 = most. The outcome variable--Parent as Teacher--was constructed by assigning scores of 1 (none/few) on variable 14-1 to teachers who were not asked the question because they indicated that no parents had visited in the past year ("no" to question 13), and scores of 1, 2, or 3 to teachers who indicated that parents had visited ("yes" to question 13).

Descriptive statistics and year-to-year correlations are reported in Appendix G. Stability was moderate: correlations between adjacent years were .37 and .31.

Minutes of Parent Time in Class. This variable was incorporated here both as an index of the teacher's success in getting parents to visit and to provide an estimate of how much parent time was available for substantive involvement in classroom activity. The measure was derived from time-sampling observation during a single class day using the CAR observation schedule, which was part of the COS. Variable scores are expressed in minutes. Parents who were paid aides were excluded from this measure.

Sample means, both overall and at the level of programs within site, were quite small, while variances were high, due to the fact that parents were present in very few classrooms on the day of observation. Apparently, the probability of encountering an unpaid parent in a classroom on a randomly selected day was low in both PDC and Comparison samples. Under these circumstances, it was not surprising that year to year correlations, for the small sample of teacher/classrooms available in more than one year, were very small and nonsignificant (Appendix G). No meaningful estimate of inter-observer reliability was possible since there were no parents present in classrooms included in the reliability study; however, we would expect inter-observer agreement about the presence of parents and the length of their stay in the classroom to be very high given the observation procedures.

Classroom Environment

Though ACYF did not indicate that it had any specific expectations regarding the general qualities of PDC classroom environments, it seemed possible that the PDC intervention might affect both the physical environment and the social-emotional climate of classrooms.

Stimulating and Attractive Physical Environment. This variable is a composite of three CEO global ratings (scales 1, 4, and 5; reproduced in Appendix G):

CE01: Degree to which materials were neatly managed and well-organized.

CE04: Degree to which the classroom was attractive/colorful.

CE05: Degree to which the classroom provided a stimulating environment for learning.

Each of these characteristics was scored on a three-point scale. The composite variable was constructed by averaging the scores for the three component variables. Though high scores on the composite variable seemed somewhat akin to "Good Housekeeping Seals of Approval," it was decided that the variable warranted inclusion in the outcome set since "image," however superficial, tends to have an important impact on visitors, parents or other members of the community, whom PDC programs were presumably committed to attract and involve.

Inter-rater agreement was high (80% to 90% perfect agreement) in the reliability study; component variables intercorrelated moderately to strongly. There was some tendency for component variable intercorrelations to increase with time. The stability of composite ratings from one year to the next was very low and failed to reach statistical significance, raising doubts about the universality of "taste" and/or the consistency of teachers' housekeeping. Descriptive statistics and correlational data are presented in Appendix G.

Supportive, Enthusiastic Climate. This general measure of classroom environment has, perhaps, more face validity than the other variable in this category. It is a composite (average) of seven global ratings (reproduced in Appendix G) made during two different days of observation. Two scales were rated twice, once each day of observation; therefore, there were only five distinct component variables:

CAR12 & FOCC21: Degree to which the teacher was affectionate and warm toward the children in her class. (Measured on two days.)

CAR15: Degree to which the teacher seemed to go out of her way to make children feel wanted and accepted.

CAR17: Degree to which children in the classroom received a great deal of encouragement from the teacher in their work.

CAR18 & FOCC20: Degree to which the teacher seemed to be enthusiastic and to really enjoy teaching. (Measured on two days.)

CAR19: Degree to which the teacher seemed to go out of her way to make children feel competent and successful.

High scores on this variable would seem very much in keeping with the values and objectives embodied in PDC. Providing a supportive, enthuiastic classroom climate was ostensibly responsive to ACYF's interests in meeting the needs of the "whole child," in accepting individual differences, in nurturing positive self-image, and in creating a setting where children would develop an interest in and commitment to schooling.

Inter-rater agreement in the reliability study was very high, ranging from 85% to 100% (for five scales) perfect agreement; independent ratings were within one scale point of each other 100% of the time for all component variables. The stability of the composite variable from one year to the next was also quite respectable; correlations ranged from .37 to .50. Component variable intercorrelations were moderate to high in 1979 (.47 to .78; median, .64) and high to very high in 1980 and 1981 (.65 to .87 and .74 to .88, respectively). Descriptive statistics and correlational data are presented in Appendix G.

Educational Management

This domain subsumes a diverse set of outcome variables, each having something important to say about teachers as managers. The first variable in the set--Maintenance of Orderly Classroom Process--is the most general index of effective management. No expectations were explicitly stated regarding general management of PDC classrooms; however, it goes without saying that a socially disorganized classroom environment would be counter-productive to PDC's objectives for children's learning and development.

The second variable--Management of Information for Individualization--is directly relevant to PDC's expressed objective of improving individualization of instruction for all children. And, in spite of considerable instrument development effort, it is the only measure of the process of individualizing instruction that the evaluation has to offer. Since at the time of PDC's conception there was little consensus regarding the operational definition of this construct, we cast our measurement net broadly in hope of measuring everyone's notion of individualization, but in doing so we seem not to have measured anyone's notion very well. The major source of information about individualization available to the evaluation was the set of post-Teacher Interview global ratings already described. Of these numerous ratings, only two met minimal criteria for inter-rater agreement. Fortunately, they happened to be items measuring the one aspect or condition of individualization about which most educators would agree--that in order to individualize instruction teachers must have specific information about the needs and status of individual learners.

While it may be generally important that teachers maintain an orderly classroom process if they are to achieve any educational goals, the research literature suggests that certain more specific aspects of classroom management are crucial to the attainment of specific academic learning objectives, measured by widely used reading and math achievement tests and commonly referred to as "basic academic skills." Though most educators, and certainly ACYF, would argue that "basic academic skills" include more than those skills required for successful performance of the microscopic tasks embodied in most achievement tests (including those tests used to measure academic outcomes in this evaluation), academic achievement narrowly defined by test performance remains the single most important indicator we have of children's social competence as students--i.e., their everyday functioning in the role of learner. Consequently, considerable effort has been made to measure aspects of the classroom process that are susceptible to management by teachers and that have been found to predict children's academic achievement.

The four remaining outcome variables in this domain are concerned with allocation of teaching-learning time in the classroom: % of Time Teaching Children; Level of Teacher/Child Involvement in Learning; % of Child Time Engaged in Any Educational Activity; % of Child Time Engaged in Learning with High Attention. Outcome variables in the fifth domain--Allocation of Learning Time--have a similar focus. However, the variables considered here measure the allocation of teacher and collective student time proportionally (disregarding the actual amounts of time involved), while variables in the Allocation of Learning Time domain estimate the actual number of minutes per/child available for and allocated to specific types of learning activity. Thus, the variables included in this domain indicate whether teachers put available personal and class time to "good" use, while Allocation of Learning Time variables represent the instructional experience of (average) individual children expressed in minutes spent doing this or that.

The past decade has witnessed a flurry of research into the relationships between the teaching-learning process and children's academic achievement. Shaping much of this research has been a concern with "learning time," and more specifically "time on task," growing out of Caroll's (1963) model of school learning. A large number of studies (reviewed in Caldwell, Huitt, & Graeber, 1982), have consistently found positive relationships between time available for learning and academic achievement, with the strength of relationship increasing as measures of time reflect more precisely the content and quality of children's activity. Other research (e.g., Rosenshine, 1981) has indicated that the quality of child learning time and academic achievement increase as teachers devote more time to instructional interaction/supervision. The outcome variables described below measure general features of the management of teaching and learning time in each classroom.

Maintenance of Orderly Classroom Process. This composite variable combines ten global ratings (reproduced in Appendix G) made on two different days of observation:

CAR11: Degree to which adults and children in the classroom had no problem making themselves heard.

FOPR16 & CAR13: Degree to which the teacher seemed to be respected and listened to by the children. (Measured on two days.)

FOPR8: Degree to which the teacher minimized time spent controlling misbehaviors.

FOPR9: Degree to which the teacher caught misbehaviors before they spread or increased in seriousness.

FOPR11: Degree to which the teacher kept classroom activities running smoothly without delays or interruptions.

FOPR14: Degree to which the teacher had no problem making herself heard.

FOPR15: Degree to which the teacher seemed calm and unruffled by classroom situations.

FOPR17: Degree to which children were cooperative in doing what was expected of them.

FOPR18: Degree to which the teacher managed the classroom well.

The composite score was constructed by averaging these ten three-point scales.

Inter-rater agreement was generally above 90%, ranging from 81% to 100% perfect agreement. Stability of composite scores was curiously very low and nonsignificant from 1979 to 1980, but quite high (r=.64, p=.001) from 1980 to 1981. Patterns of component variable intercorrelations were quite similar across the three years; correlations ranged from .17 to .97 with median values for all three years greater than .58.

Management of Information for Individualization. This variable is a composite of two three-point rating scales from the Teacher Interview (scales 8 and 9; reproduced in Appendix G):

- Degree to which teacher maintains specific and comprehensive records on each child that contain a variety of information, such as observations, test results, and work specimens.

- Degree to which teacher appears to have specific knowledge of individual children's strengths, needs, problems, and interests related to language arts and math.

The component variables were averaged to form the composite.

Inter-rater agreement was better than 80% based upon limited data obtained during 1980 interviewer training; stability of the composite rating across one-year intervals varied from very low and nonsignificant (1979-80) to moderate and highly significant (r=.32, p=.008 for 1980-81). Descriptive statistics and correlational data are presented in Appendix G.

Proportion of Teacher Time Spent Teaching. This variable was derived from the CAR time-sampling observation data by dividing the number of minutes the teacher spent teaching children by the number of minutes in the class day (CAR Class-Level Variables [CLV] 2 and 3; described in Appendix G). Interobserver agreement (intraclass correlation coefficient) was high (.92 and .82) for first-order observations made during 5-minute intervals; stability of the outcome variable was low and nonsignificant for 1979-80 but moderate (r=.46, p=.009) for 1980-81. Descriptive statistics and correlational data are presented in Appendix G.

Level of Teacher/Child Involvement in Learning. This variable was constructed by averaging scores for three COS three-point rating scales (reproduced in Appendix G):

FOCC3:	Degree to which children seemed interested and attentive to the learning activities provided.
FOCC14 & CAR14	Degree to which there was interest and involvement in learning activities on the part of both adults and children in the classroom. (Measured on two days.)

This composite variable provides a global estimate of the quality of learning time in each classroom--specifically, the global *engagement rate* for teachers and children. Inter-rater agreement in the 1981 reliability study was high (85% to 100% perfect agreement); stability of composite variable scores from one year to the next was moderate to high (r=.31, p=.038 for 1979-80; r=.51, p=.004 for 1980-81). Descriptive statistics and correlational data are presented in Appendix G.

Proportion of Child Time Engaged in Any Educational Activity. The denominator of this variable restricts available child time to *total child time in the classroom*; it specifically excludes time spent by individual children or the entire class outside the classroom, be that time spent in recess, eating lunch, going to the toilet, attending assemblies, or whatever. The numerator includes the *total number of minutes spent by all children in the class nominally engaged in any purposive educational activity*, including not only specific language arts and math instruction but also social studies, science, music and movement, arts and crafts, second language instruction, and so forth. The variable was derived from CAR time-sampling observation data (CAR CLVs 11 and 21; described in Appendix G). Interobserver reliability for first-order observations during 5-minute intervals was estimated by intraclass correlation to be .72 for one component (CAR page level variable [PLV] 20) and by percent agreement to be 82% perfect agreement for the other component (CAR PLV 9; see Appendix G for discussion of reliability estimation). Stability was low from 1979-80 but high (r=.54, p=.002) for 1980-81. Descriptive statistics and correlational data are presented in Appendix G.

Proportion of Child Time Engaged in Learning with High Attention. As with Level of Teacher/Child Involvement in Learning, this variable measures the quality of learning time; however, estimates of engagement rate are restricted to children in the classroom and to specific language arts and math activities. The variable was derived from the CAR time-sampling observation data by dividing the *total number of child minutes devoted to language arts and math activities with high attention by the total number of child minutes devoted to language arts and math activities* (CAR CLVs 14-17; described in Appendix G). Interobserver reliability (intraclass correlation coefficient) for first-order observations during 5-minute intervals ranged from .72 to .89 for the four component variables; when coefficients were computed for 10-minute intervals, reliability increased, ranging from .80 to .92. Stability of the outcome variable was low and nonsignificant for 1979-80, but moderate (r=.40; p=.027) for 1980-81. Descriptive statistics and correlational data are presented in Appendix G.

PDC-Encouraged Instructional Content

By way of project staffing and formal training implementation of the PDC Guidelines was intended to affect elementary school curricula in three areas traditionally emphasized in Head Start programs--multicultural education, health and nutrition education, and utilization of community resources. The frequency with which teachers used materials/activities to further multicultural or health and nutrition education, and used materials or people from the community in classroom activities was measured by three Teacher Interview questions (9 i, o, and l, respectively; reproduced in Appendix G). Frequency of use was measured on a six-point scale: 1 = never; 2 = about once or twice this year; 3 = about every other month; 4 = about once a month; 5 = two to three times a month; 6 = once a week or more. Each of these interview variables was treated as a distinct outcome measure:

- Multicultural Instruction
- Health/Nutrition Instruction
- Use of Community Resources

Stability of variable scores over time was moderate to high for all three variables, ranging from .31 to .59. Descriptive statistics and correlational data are presented in Appendix G.

Allocation of Learning Time

The variables in this category provide estimates of how the hypothetical "average child" in each classroom spent his/her time during the day of observation. Theoretically, of the various teacher/classroom outcomes measured, these variables should be the best predictors of children's academic achievement. Though our measures of quality of learning time were not as precise as those used in some basic research (e.g., Academic Learning Time; see, Fisher, Berliner, Filby, Marliane, Cahen, & Dishaw, 1981), analogous measures have been demonstrated to predict achievement (Caldwell, et al., 1982) when obtained under suitable conditions. What constitutes sufficiently suitable conditions is difficult to say. However, if one wished to predict achievement outcomes for individual children, it would seem most desirable to observe individual children and to observe them on several occasions in order to obtain reliable characterizations of their learning experience. This was not done in the PDC evaluation: observations of learning time were made on a single day for all groups of children in the class ("groups" ranging in size from the whole class to one or two children working independently); and in constructing the outcome measures considered here, we aggregated characterizations of group behavior made during 5-minute intervals to the level of class and class day, then estimated the behavior of a hypothetical

average child for each classroom. In spite of these limitations, exploratory analyses of the relationships between the classroom experiences of hypothetical average children and the achievement of actual individual children enrolled in the same classrooms revealed frequently significant correlations in the expected direction, suggesting that the outcome variables described below do provide useful characterizations of the teaching-learning process at the level of individual classrooms. And it seemed likely that characterizations of groups of classrooms (e.g., within programs within sites), though abstract, would be even more reliable and valid. Descriptive statistics and correlational data for these variables are presented in Appendix G.

Minutes/Child Available for Learning. This is an estimate of the *number of minutes that the average child was present in the classroom*--the maximum time available for learning. It was derived from CAR time-sampling observations (CAR PLV 9; described in Appendix G). Interobserver agreement for first-order observations during 5-minute intervals was 82% (PLV 9); for 10-minute intervals, agreement was 88%. Stability was very low across the three years during which data were collected. Whether observed changes in classroom characterizations were the product of short-term or long-term fluctuations in the teaching-learning process cannot be determined.

Minutes/Child Allotted to Specific Academic Activities: *the number of minutes that the average child spent occupied in activities in mathematics, reading or other language arts pursuits, science or social science.* The variable was derived from CAR time-sampling observations (CAR PLVs 13, 15 and 18; described in Appendix G). Interobserver agreement was .72 for "other academic" categories (PLV 18)--science, social science--and .89 and .91 for reading/language arts and mathematics, respectively. PLV 18 showed a decline in interobserver agreement when 10-minute intervals were considered, so that average percent perfect agreement was computed as a check: it was 96%. Stability was moderate between adjacent years (.35 and .36); paradoxically, it was high (.82) between 1979 and 1981 for the tiny sample (n=10) that was observed on both of those occasions.

Minutes/Child Allotted to Mathematics: *the number of minutes that the average child spent in math activities* (CAR PLV 13). As reported above, interobserver agreement (intraclass correlation) was .91 for observations during 5-minute intervals, and .94 when 10-minute intervals were used. Stability over time was very low; none of the correlations between different years reached statistical significance.

Minutes/Child Allotted to Reading/Language Arts: *the number of minutes that the average child spent in English reading and writing activities* (CAR PLV 15). Intraclass correlations were .89 for 5-minute intervals and .92 for 10-minute intervals in the 1981 reliability study. Stability over time was moderate (.37) between 1980 and 1981, but too low to reach statistical significance between 1979 and 1980.

Minutes/Child Engaged in Mathematics with High Attention: *the number of minutes the average child spent engaged in mathematics activities during which attention was coded "high"* (PLV 14; described in Appendix G). Interobserver agreement was .72 for 5-minute intervals, and rose to .80 when 10-minute intervals were considered. Stability across time for these variables was low, which is not surprising since stability for Minutes/Child Allotted to Mathematics was also low.

Minutes/Child Engaged in Reading/Language Arts with High Attention: *the number of minutes the average engaged in English literacy activities and in which attention was coded as "high"* (PLV 16; described in Appendix G). Interobserver agreement was .75 for 5-minute intervals of observation, rising to .81 when 10-minute intervals were considered. The correlation between 1980 and 1981 scores for this variable was .36, judged to be a sign of moderate stability; the correlation between 1979 and other years did not reach statistical significance.

Overview of Teacher/Classroom Outcome Measures

With the exception of two domains--Promotion of Parent Involvement and PDC-Encouraged Instructional Content--measurement of teacher/classroom outcomes was guided not so much by ACYF's expressed PDC goals as by current interests and knowledge in the field of educational research. The instruments developed to measure outcomes in the Educational Management and Learning Time domains were intended to provide information about aspects of classroom process and children's learning experience that had been demonstrated in other research to predict children's academic success. The fact that these measurement procedures were necessarily less refined than those used in some, more basic research on the teaching-learning process severely limits our ability to model teaching-learning relationships; however, we are fairly confident that the data collected are sufficient to reveal major differences in classroom process and learning experience between site-level and aggregate PDC and Comparison groups if such differences occurred.

As for measuring teachers' efforts to further nonacademic objectives such as children's social-emotional development, we did not. The combined efforts of evaluation staff, the Advisory Panel, and ACYF officers were not sufficient to either identify or develop satisfactory measures of such outcomes.

ANALYTIC DESIGN

The basic methods used to analyze teacher/classroom outcomes and other quantitative data are described in detail in Chapter II--Data Analysis Methods Used in the Final Evaluation. First-order findings from Designs 1 and 2 are presented in Appendix I, as are observed, predicted, and ANCOVA-adjusted means for each outcome by group by site and by aggregate PDC and Comparison groups. A synthesis of Design 1 and 2 findings is presented in this chapter.

Information about teachers and classrooms was collected in spring 1979, 1980, and 1981 through interviews and classroom observations. About one third of teacher/classroom outcomes were measured by interview; the remaining two thirds, by observation. Observations (using the COS) were made only of classrooms attended by children in the evaluation cohort; however, interviews were conducted with randomly selected teachers at each grade level at each PDC school and selected Comparison schools in addition to teachers of children in the cohort. Thus, the interview sample was substantially larger than, and quite different from, the observation sample. As mentioned earlier in this chapter, a decision was made to restrict the analytic sample to teachers of and classrooms attended by children in the longitudinal analytic sample, a sample equivalent to the observation sample minus a few classrooms attended only by children who were later lost to attrition. Restricting the analytic sample in this way greatly facilitated integration and interpretation of findings since the results from analyses of interview and observation data pertained to the same teachers and classrooms. Moreover, by restricting the sample to teachers and classrooms of children in the longitudinal analytic sample we increased the likelihood that teacher/classroom outcomes could be interpreted as linking institutional with child impacts.

The specific analytic design for teacher/classroom outcomes was, of necessity, rather different from the designs for parent and child outcomes. While the group-by-site samples of parents and children were small (averaging 16 subjects), by comparison the annual group-by-site samples of teachers and classrooms were minute. The average number of PDC classrooms per site per year in the analytic sample was less than 5 (ranging from 1 to 8); the average number of Comparison classrooms, less than 9 (ranging from 3 to 20). The power of statistical tests applied to annual site-level samples would have been totally unacceptable at all but one or two sites.

In order to increase the sizes of site-level samples, and thereby the power of tests performed at the site level, we explored the feasibility of pooling teacher/classroom data across years (1979-1981) within groups within sites, and ultimately decided to analyze data pooled in this manner. This approach meant that statistical tests performed on teacher/classroom data were of roughly the same power as tests of parent and child outcomes, and occasionally more powerful. However, pooling data collected at three time points posed a new problem, which had to be addressed before data analysis

could proceed sensibly and interpretably. Some teachers had been interviewed and observed on more than one occasion--36 in both 1979 and 1980, 26 in both 1980 and 1981, and 4 in all three years. This occurred for one of two reasons: either a cohort child had been retained in grade (about 60% of cases) or a teacher had changed the grade level at which s/he taught (about 40% of cases). The inclusion of repeated measures for teachers of children who had been retained in grade resulted in over-representation of lower grade classrooms, while inclusion of any repeated measures resulted in over-representation of individual teachers.

Our interest was in PDC's impact on teachers in general, not on a subsample of teachers who happen to have been interviewed and observed on more than one occasion. Yet at some sites, particularly in small PDC samples, the repeated measures of a few teachers constituted up to one half of the available data. Furthermore, examination of outcomes for these teachers indicated that they were sometimes very different from outcomes for other teachers in the sample who had not been measured repeatedly. Rather than disregard a situation that would further cloud the meaning of evaluation findings and rather than make an already complicated and precarious analytic design even more so by attempting to accomodate "partially repeated measures", we opted for random deletion of repeated measures so that each teacher would be represented by data collected at only one point in time. This reduced the total number of cases in the pooled data set from 392 to 319 (cf., Table V-1), eliminating over-representation of individual teachers and of lower grade classrooms without reducing the sizes of group-by-site samples below levels prevailing in analyses of parent and child outcomes.

The distribution of teachers/classrooms by grade levels in the final analytic sample is illustrated in Table V-2 for each year and for the total analytic sample. The information in that table reveals yet another problem that had to be addressed before analyzing teacher/classroom outcome data. There was a substantial number of mixed-grade classrooms in the analytic sample. Such classrooms were as likely to occur in the Comparison as in the PDC group; however, mixed-grade classrooms did not usually occur in equal proportions in PDC and Comparison groups at particular sites, nor were particular mixes equally represented in group-by-site samples. Often related to such differences in "mixedness" were differences between PDC and Comparison site-level samples with respect to the proportions of single-grade classes at each grade level. That this situation represented a potential problem was clearly indicated by preliminary analyses which revealed both strong grade-level trends and large differences between mixed-grade and single-grade classrooms on many outcome variables. Thus, group differences in grade mix/level were likely to influence, and potentially bias, mean outcome levels at most sites.

Whether the effects of group differences in grade mix/level would be "biasing" (and therefore, whether such differences should be controlled in the analytic design) was difficult to decide. If the presence of mixed-grade classrooms in a particular site-level sample were the result of PDC program policy, for example, then it might be argued that group

Table IV-2

Distribution of Classrooms by Grade Levels
for Each Year of Measurement and the Pooled Analytic Sample

	Spring 1979 N	Spring 1980 N	Spring 1981 N	Across All Three Years N
Kindergarten	11	-	-	11
Grade 1	81	17	-	98
Grade 2	-	74	26	99
Grade 3	-	-	81	82
Mixed-Grade--Total	14	11	4	29
K & 1	3	-	-	3
K - 1 - 2	1	-	-	1
1 & 2	7	6	1	14
1 - 2 - 3	2	-	-	2
1 - 2 - 3 - 4	1	-	-	1
2 - 3	-	5	1	6
3 - 4	-	-	2	2
Total Number of Teachers/Classrooms	106	102	111	319

differences with respect to "mixedness" were legitimate treatment effects that might influence, but would not bias, teacher/classroom outcomes. With regard to the proportions of single-grade classrooms at each grade level, we could imagine no situation, *in the absence of mixed-grade classrooms*, in which group differences in proportions would have resulted from program differences. Examination of the distributions of mixed-grade classrooms across schools within groups at each site revealed only one site where it appeared that a mixed-grade policy might have produced group differences in mixedness--all PDC children in the evaluation cohort had been assigned to mixed-grade classes for their first-grade year. Upon further investigation, however, it was discovered that the large PDC school at that site had exclusively first-grade classrooms as well, but that by chance none of the children in the evaluation cohort was assigned to a single-grade class for first grade. Thus, insofar as we could determine, group differences in the proportions and types of mixed-grade classrooms represented in the analytic sample did not result from systematic differences between PDC and school district programs/policies, but from sampling accidents and/or school-level decisions that often fluctuated from year to year in response to changes in the school's student population.

In the absence of any evidence that group differences in grade mix/level were a function of systematic differences in local programs, we concluded that the presence of rather substantial differences in grade mix/level at most sites would bias teacher/classroom outcomes. Therefore, an effort was made to control for these differences by incorporating measures of grade mix and level as covariates in the analytic design. The alternative of blocking on class type--comparing grade 2 PDC classes with grade 2 Comparison classes, K-1-2 PDC classes with K-1-2 Comparison classes, and so forth--was entertained only momentarily. Though such an approach has obvious advantages, it was impractical even for comparisons of aggregate PDC and Comparison samples given the very small number of most types of mixed-grade classes. At the site level, such an approach was out of the question.

Yet another potential source of bias was considered in the course of developing the specific analytic design for teacher/classroom outcomes--group differences in teacher characteristics unaffected by the intervention. Such characteristics included sex, ethnicity, age, previous education, previous teaching experience, and so forth. Preliminary analyses indicated that PDC and Comparison teachers frequently differed with respect to one or more of these characteristics at the site level. However, only one measured characteristic--previous teaching experience--reliably predicted outcomes, and even in that case predictions were weak. decision to include or exclude this variable as a covariate hinged on whether or not it was truly "unaffected by the intervention." Though obviously PDC could not have altered the previous teaching experience of an individual teacher, local PDC programs might have influenced the composition of teacher samples with respect to previous teaching experience. Such influence might have resulted from deliberate personnel decisions that consituted part of the change strategy and/or from the inadvertent

creation of pressures within the institutional context for self-selection in or out of the program.

At four of five sites where significant group differences were found, PDC teachers were less experienced on the average than Comparison teachers. We have no knowledge of explicit personnel policies within local schools and programs that would explain these findings; however, we have extensive anecdotal data from site visits that suggest there were selection pressures operating in at least some PDC schools/programs which might explain the differences observed. Specifically, it was often reported that older, more experienced teachers were reluctant to adopt the PDC innovation. Many of these teachers were reported to be seeking and obtaining transfers to other schools where they could run their classrooms as they wanted; others were said to have had transfers sought for them by building principals seeking to reduce conflict among staff. Teacher turnover was extremely high in some PDC schools, and there appears to have been a downward trend in teachers' previous teaching experience even within our small analytic sample from 1979 to 1981 at a majority of sites. Even at one site where virtually the entire staff of youngish PDC teachers were laid off and replaced by teachers with greater seniority, the average number of years of teaching experience at the same or an adjacent grade level was significantly lower for PDC than Comparison teachers--due apparently to the reassignment of senior teachers to whatever openings were available without regard to their own grade-level preferences or experience. Particularly in the case of Educational Management and Learning Time outcomes, prior experience at the same grade level was found to be positively correlated, albeit not very strongly (less than .25).

Given that group differences in teacher characteristics seemed more a product of the PDC intervention than of pre-existing differences and that, in any case, teacher background (as measured) did not strongly predict outcomes, we decided not to incorporate background variables as covariates. The covariates ultimately included in the analytic design are described below.

Covariates Incorporated in the Analytic Design

All of the covariates used in analyzing teacher/classroom outcomes were intended to control for group differences, particularly at the site level, in grade mix/level. Given sometimes dramatic differences between site-level PDC and Comparison groups in the proportions of mixed- versus single-grade classes and/or the proportions of classrooms at different grade levels, we sought to represent these differences as completely as possible in the covariate set. Toward that end three approaches were taken.

First, a single dummy variable was created to represent "mixedness"--mixed-grade/single-grade. Second, a set of seven dummy variables was

created to represent the grade level at which each class was "centered." These variables were derived from the following eight-point grade continuum:

- Kindergarten
- Kindergarten-First Grade Mix
- First Grade, Kindergarten-First-Second Grade Mix
- First-Second Grade Mix
- Second Grade, First-Second-Third Grade Mix
- Second-Third Grade Mix, First-Second-Third-Fourth Grade Mix
- Third Grade
- Third-Fourth Grade Mix

Third, having long ago despaired of estimating program-related changes in teacher/classroom outcomes over time, we decided to incorporate *year of measurement*--represented by two dummy variables--to control for group differences in the distribution of class types by year. Together, these ten covariates accounted for as much variance in teacher/classroom outcomes as we seemed able to explain having experimented with myriad ways of representing grade mix and level. The covariates were particularly effective in removing variance from outcomes in the Learning Time domain--accounting for 10 to 39% of outcome variance in both Design 1 and Design 2 (see Appendix I). Since preliminary analyses (disregarding group) indicated that Learning Time outcomes were most influenced by grade level and mixedness, the behavior of the covariate set was encouraging. However, given the substantial group differences in class types at some sites, we still harbor doubts about the ultimate effectiveness of the analytic design in eliminating bias from teacher/classroom outcomes at the site level.

FINDINGS

Methods of analysis are described and the presentation of findings is explained in detail in Chapter II--Data Analysis Methods Used in the Final Evaluation.

Net effects of consequence are summarized graphically in Figure V-1. All estimates of net observed and net predicted effects are reported in Table V-3; net effects judged consequential are highlighted. The reader is reminded that primary estimates of net effects were based on pooled *probable* and *possible (??)* findings, with the cutoff for consequentiality set at .25. Site-level and aggregate findings from the synthesis of Designs 1 and 2 are presented in Tables V-4 through V-6 for *probable* effects, *probable and possible (??)* effects, and *probable, possible (?), and less possible (??)* effects, respectively. These tables show (at three levels of probability) the patterns of findings across time and across variables within outcome domains upon which estimates of net effect were based for each site and the total sample. Primary estimates of net effects and judgements of consequentiality were based on the patterns shown in Table V-5 for *probable and possible* effects. Detailed results of all Design 1 and 2 analyses are reported in Appendix I.

Summary of Findings for Teacher/Classroom Outcomes

This chapter addresses the following question:

> *Did PDC teachers exhibit attitudes, knowledge, and/or behaviors toward children that were different from those of non-PDC teachers and, at the same time, compatible with the intentions of the Guidelines?*

In answer to the question: *yes* and *no*, both at the site level and in the aggregate. Findings are briefly summarized by domain below and in Figure V-1; they are discussed at some length in the next section.

Promotion of Parent Involvement

PDC teachers appear to have done more to promote parent involvement in the classroom than teachers in Comparison schools at six of ten sites--Sites 2, 3, 4, 5, 6, and 9. No Comparison-favoring net effects (consequential or otherwise) were found. A rather strong (.75) net positive effect was found over all sites.

Figure V-1

Teacher/Classroom Outcomes
Consequential Net Effects by Domain and Site

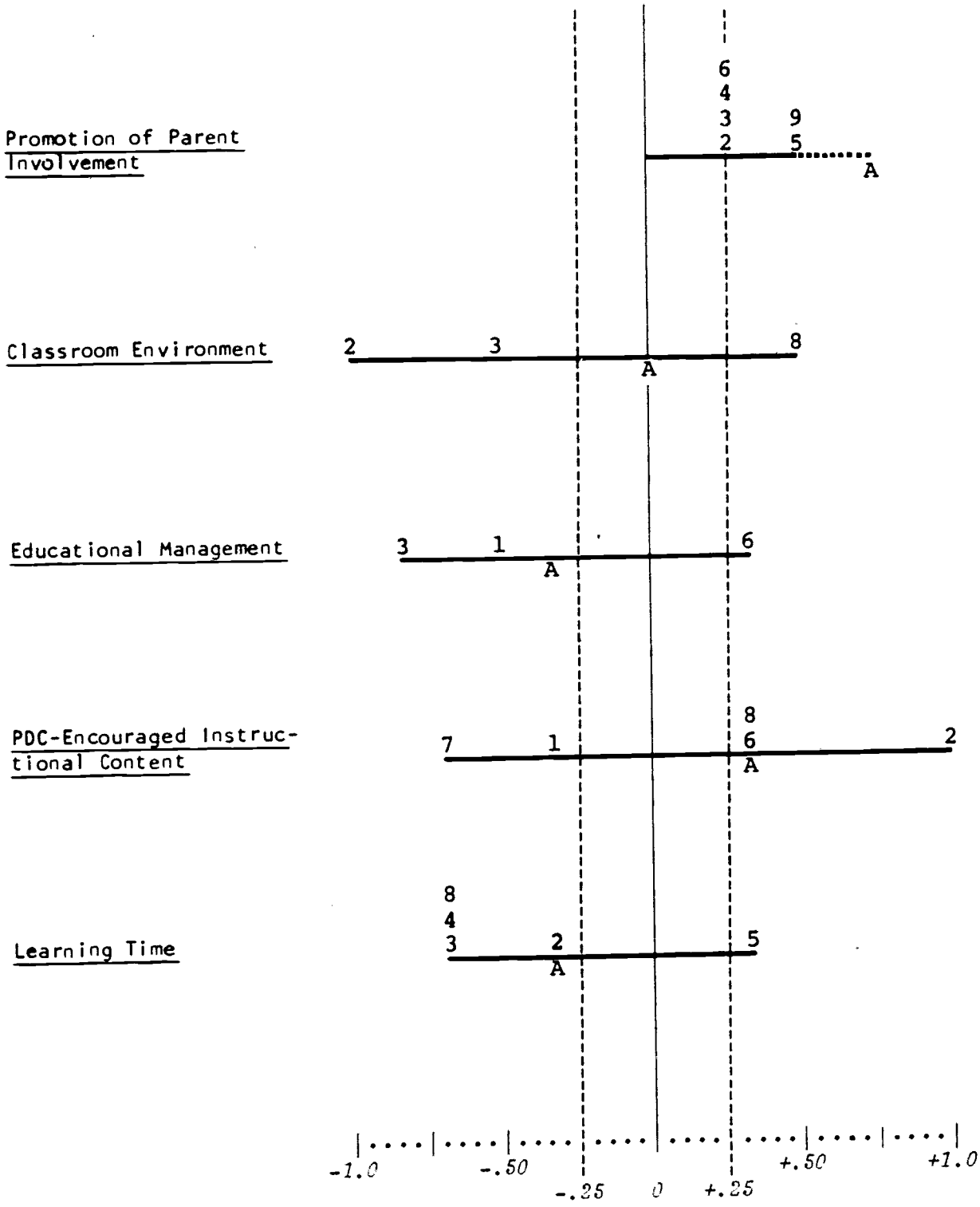


Table V-3

Teacher/Classroom Outcomes
Summary of Net Observed Effects and Predictions by Site
and for Aggregate PDC and Comparison Samples at Three Levels of Probability

		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Agg
Promotion of Parent Involvement	Probable	.00	.00	.25	.00	.50	.00	.00	-.25	.50	.00	.50
	Prob/?	.00	.25	.25	.25	.50	.25	.00	.00	.50	.00	.75
	P/?/??	.00	.25	.25	.25	.50	.25	.00	.25	.50	.00	.75
	Predicted	.00	.00	.00	.00	.00	.00	.00	-.25	.00	.00	.00
Classroom Environment	Probable	.00	.00	-.50	.00	.00	.00	.00	-.50	.00	.00	.00
	Prob/?	.00	-.50	-.50	.00	.00	.00	.00	.50	.00	.00	.00
	P/?/??	.00	-1.0	-.50	.00	.00	.00	.00	.50	.00	.00	.00
	Predicted	.00	1.0	.00	.00	.00	.00	.00	.00	.00	.00	.00
Educational Management	Probable	-.33	.00	-.83	-.17	.00	.17	.00	.00	.00	.17	.00
	Prob/?	-.50	.00	-.83	-.17	.00	.33	.00	.17	.00	.17	-.33
	P/?/??	-.50	-.50	-.83	-.17	.00	.33	.00	.33	.00	.17	-.33
	Predicted	.00	.50	.00	.00	.00	.00	.00	.00	.00	.00	.00
PDC-Encouraged Instructional Content	Probable	.00	.00	.00	.00	.00	.33	-.33	.33	.00	.00	.00
	Prob/?	-.33	1.0	.00	.00	.00	.33	-.33	.33	.00	.00	.33
	P/?/??	-.33	1.0	.00	.33	.00	.33	-.67	.33	.00	.00	.33
	Predicted	.00	.00	.00	-.33	.00	.00	.33	.00	.00	.00	.00
Learning Time	Probable	.17	-.17	-.67	-.33	.33	.00	.00	-.50	.00	-.17	-.33
	Prob/?	.17	-.33	-.67	-.67	.33	.17	.17	-.67	-.17	-.17	-.33
	P/?/??	.17	-.33	-.67	-.67	.33	.17	.17	-.67	-.17	-.17	-.33
	Predicted	.00	.17	.00	.00	.00	.00	-.17	.00	.00	.00	.00

Table V-4

Summary of Probable Effects for Teacher/Classroom Outcomes
Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site and the Aggregate Sample

Outcome Domains and Measures	Sites										Agg
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
Promotion of Parent Involvement											
Attitude Toward Parent Involvement	0	0	0	0	+	0	0	0	+	0	+
Amount of Home Visiting	0	0	0	0	+	0	0	-	0	0	0
Parents as Teachers	0	0	0	0	0	0	0	0	+	0	+
Minutes of Parent Time in Class	0	0	+	0	0	0	0	0	0	0	0
Classroom Environment											
Stimulating & Attractive Environment	0	0	0	0	0	0	0	+	0	0	0
Supportive & Enthusiastic Climate	0	0	-	0	0	0	0	0	0	0	0
Educational Management											
Maintenance of Orderly Classroom Process	-	0	-	0	0	0	0	0	0	0	0
Management of Information for Individualization	-	0	0	0	0	0	0	0	0	+	0
% of Time Teaching Children	0	0	-	0	0	0	0	0	0	0	0
Level of Teacher & Child Involvement in Learning	0	0	-	0	0	0	0	0	0	0	0
% of Child Time Engaged in Any Educational Activity	0	0	-	-	0	+	0	0	0	0	0
% of Child Time Engaged in Learning with High Attention	0	0	-	0	0	0	0	0	0	0	0
PDC-Encouraged Instructional Content											
Multicultural Instruction	0	0	0	0	0	0	0	+	0	0	0
Health/Nutrition Instruction	0	0	-	0	0	+	-	0	0	0	0
Use of Community Resources	0	0	+	0	0	0	0	0	0	0	0
Learning Time											
Minutes/Child Available for Learning	0	0	0	0	0	0	0	-	0	-	0
Minutes/Child Allotted to Specific Academic Activity	0	0	-	-	+	0	0	-	0	0	0
Minutes/Child Allotted to Math	+	0	0	0	0	0	0	0	0	0	0
Minutes/Child Allotted to Reading/Language Arts	0	-	-	-	+	0	0	-	0	0	-
Minutes/Child: Math with High Attention	0	0	-	0	0	0	0	0	0	0	0
Minutes/Child: Reading/Language Arts with High Attention	0	0	-	0	0	0	0	0	0	0	-

Table V-5

Summary of Probable and Possible (?) Effects for Teacher/Classroom Outcomes
Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site and the Aggregate Sample

Outcome Domains and Measures	Sites										Agg
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
Promotion of Parent Involvement											
Attitude Toward Parent Involvement	0	0	0	0	+	0	0	0	+	0	+
Amount of Home Visiting	0	+	0	0	+	0	0	-	0	0	0
Parents as Teachers	0	0	0	0	0	+	0	+	+	0	+
Minutes of Parent Time in Class	0	0	+	+	0	0	0	0	0	0	+
Classroom Environment											
Stimulating & Attractive Environment	0	-	0	0	0	0	0	+	0	0	0
Supportive & Enthusiastic Climate	0	0	-	0	0	0	0	0	0	0	0
Educational Management											
Maintenance of Orderly Classroom Process	-	0	-	0	0	0	0	0	0	0	-
Management of Information for Individualization	-	0	0	0	0	0	0	0	0	+	0
% of Time Teaching Children	0	0	-	0	0	+	0	0	0	0	0
Level of Teacher & Child Involvement in Learning	-	0	-	0	0	0	0	+	0	0	-
% of Child Time Engaged in Any Educational Activity	0	0	-	-	0	+	0	0	0	0	0
% of Child Time Engaged in Learning with High Attention	0	0	-	0	0	0	0	0	0	0	0
PDC-Encouraged Instructional Content											
Multicultural Instruction	0	+	0	0	0	0	0	+	0	0	0
Health/Nutrition Instruction	0	+	-	0	0	+	-	0	0	0	0
Use of Community Resources	-	+	+	0	0	0	0	0	0	0	+
Learning Time											
Minutes/Child Available for Learning	0	0	0	-	0	0	0	-	0	-	0
Minutes/Child Allotted to Specific Academic Activity	0	0	-	-	+	+	0	-	0	0	0
Minutes/Child Allotted to Math	+	0	0	0	0	0	0	-	0	0	0
Minutes/Child Allotted to Reading/Language Arts	0	-	-	-	+	0	0	-	0	0	-
Minutes/Child: Math with High Attention	0	0	-	0	0	0	+	0	0	0	0
Minutes/Child: Reading/Language Arts with High Attention	0	-	-	-	0	0	0	0	-	0	-

110

Table V-6

Summary of Probable, Possible (?), and Less Possible (??) Effects for Teacher/Classroom Outcomes Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site and the Aggregate Sample

Outcome Domains and Measures	Sites										Agg
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
Promotion of Parent Involvement											
Attitude Toward Parent Involvement	0	-	0	0	+	0	0	0	+	0	+
Amount of Home Visiting	0	+	0	0	+	0	0	-	0	0	0
Parents as Teachers	0	0	0	0	0	+	0	+	+	0	+
Minutes of Parent Time in Class	0	+	+	+	0	0	0	+	0	0	+
Classroom Environment											
Stimulating & Attractive Environment	0	-	0	0	0	0	0	+	0	0	0
Supportive & Enthusiastic Climate	0	-	-	0	0	0	0	0	0	0	0
Educational Management											
Maintenance of Orderly Classroom Process	-	-	-	0	0	0	0	0	0	0	-
Management of Information for Individualization	-	0	0	0	0	0	0	+	0	+	0
% of Time Teaching Children	0	0	-	0	0	+	0	0	0	0	0
Level of Teacher & Child Involvement in Learning	-	-	-	0	0	0	0	+	0	0	-
% of Child Time Engaged in Any Educational Activity	0	0	-	-	0	+	0	0	0	0	0
% of Child Time Engaged in Learning with High Attention	0	-	-	0	0	0	0	0	0	0	0
PDC-Encouraged Instructional Content											
Multicultural Instruction	0	+	0	0	0	0	-	+	0	0	0
Health/Nutrition Instruction	0	+	-	0	0	+	-	0	0	0	0
Use of Community Resources	-	+	+	+	0	0	0	0	0	0	+
Learning Time											
Minutes/Child Available for Learning	0	0	0	-	0	0	0	-	0	-	0
Minutes/Child Allotted to Specific Academic Activity	0	0	-	-	+	+	0	-	0	0	0
Minutes/Child Allotted to Math	+	0	0	0	0	0	0	-	0	0	0
Minutes/Child Allotted to Reading/Language Arts	0	-	-	-	+	0	0	-	0	0	-
Minutes/Child: Math with High Attention	0	0	-	0	0	0	+	0	0	0	0
Minutes/Child: Reading/Language Arts with High Attention	0	-	-	-	0	0	0	0	-	0	-

Classroom Environment

Findings in this domain were mixed. Comparison-favoring (net negative) effects were indicated at Sites 2 and 3; a PDC-favoring effect, at Site 8. There was absolutely no indication of effect over all sites.

Educational Management

PDC teachers exhibited less management skill than teachers in local Comparison schools at two sites--Sites 1 and 3--and more skill at one site--Site 6. Over all sites Comparison teachers appeared to be more skillful managers of the educational process, on average, than PDC teachers.

PDC-Encouraged Instructional Content

At Sites 2, 6, and 8 PDC teachers gave more emphasis than Comparison teachers to instructional content specifically encouraged by the Guidelines--multicultural activities, health/nutrition instruction, and utilization of community resources in the classroom. However, Comparison teachers gave more emphasis to these content areas at two sites--Sites 1 and 7. Over all sites the net effect favored PDC (+.33).

Learning Time

The average PDC child spent less time engaged in learning activities with specific academic focus than the average Comparison child at four sites--Sites 2, 3, 4, and 8. A net positive effect was found in one site--Site 5. Over all sites the net effect favored Comparison classrooms (-.33).

Discussion

Although PDC Guidelines did not require specific teacher/classroom outcomes, they did require implementation of various institutional mechanisms intended to promote outcomes in two domains--Promotion of Parent Involvement and PDC-Encouraged Instructional Content. Moreover, ACYF expressed fairly explicit objectives with respect to both domains in hope that local projects would develop their own programs for realizing these objectives within the institutional framework created by Guideline implementation. The three remaining outcome domains--Classroom Environment, Educational Management, and Learning Time--reflect not so much specific objectives of PDC as they do general standards of educational quality and assumed prerequisites for effective instruction. Findings for these two groups of domains will be discussed separately below.

PDC-Specific Outcome Domains

Promotion of Parent Involvement. The relationships between institutional outcomes and parent outcomes were examined in the preceding chapter. We concluded that implementation of institutional features prescribed by the Parent Involvement component of the PDC Guidelines, even when this resulted in appreciable PDC-favoring differences in provisions for parent involvement, did not ensure higher levels of involvement in school affairs or more effective out-of-school support for children's learning by PDC parents. It was noted, however, that parent outcome data pertained only to parents of children in the longitudinal analytic sample, an extremely small and potentially unrepresentative sample of all parents of children participating in PDC over the course of the project. Possibly PDC had more positive impact on parents in general than on parents of those relatively few children included in the longitudinal evaluation. Data collected using the Teacher Interview and Classroom Observation System pertain to a much larger sample of parents--specifically, parents of all children enrolled in the same classrooms as children in the longitudinal analytic sample. Since about 70% of classrooms contained no more than two children from the analytic sample, inclusion of parents of other children in these classrooms dramatically expanded the universe of parents.

The domain Promotion of Parent Involvement incorporates four variables. The first--Attitude toward Parent Involvement--indexes teachers' views of the potential benefits of involving parents in the classroom process. The second--Amount of Home Visiting--measures teachers' overtures to parents. The two remaining variables--Parents as Teachers and Minutes of Parent Time in Class--measure actual involvement of parents in the classroom. Though intercorrelations (ranging from .00 to .18) did not indicate that these variables formed a tightly knit set measuring a unidimensional construct, their conceptual fit, as measures of different facets of a multidimensional construct, is arguably better. And it seemed reasonable to assume that sites with PDC-favoring differences in institutional provisions for parent involvement might also exhibit net positive effects in Promotion of Parent Involvement.

Though we anticipated that findings for parents in general (represented by data analyzed in this chapter) might be different from findings for parents of children in the evaluation sample (data analyzed in Chapter IV), we hoped to find some evidence of triangulation among institutional (Chapter III), parent (Chapter IV), and teacher/classroom outcomes. Findings pertaining to parents from analyses of all three outcome sets are summarized below:

	Sites										Agg
	1	2	3	4	5	6	7	8	9	10	
Predictions from Institutional Outcomes: Parent Involvement component (Chapter III)	+	+	0	0	+	+	+	0	+	0	+[1]
Findings from Parent Outcomes: Parent Involvement and Parent as Educator domains (Chapter IV)[2]	+	0	-	0	0	-	+	0	0	+	0
Findings from Teacher/Classroom Outcomes: Promotion of Parent Involvement domain	0	+	+	+	+	+	0	0	+	0	+

[1]Median site-level effect.

[2]Synthesis of effects from two domains: +/+ = +; +/0 = +; 0/0 = 0; 0/- = -.

Predictions from institutional outcomes to Promotion of Parent Involvement at the teacher/classroom level were correct for six of ten sites, somewhat better than predictions of Parent Involvement and Parent as Educator outcomes for parents of children in the evaluation cohort (correct at only four sites). The absence of grossly contradictory findings, represented by net negative effects, for the teacher/classroom domain makes a success rate of six out of ten seem a bit more impressive, even though net positive effects were found at only four of six sites where they were expected. Predictions of aggregate effects were confirmed for the relevant teacher/classroom outcome domain, but not for parent outcome domains. Finally, predictions aside, it should be noted that consequential net positive effects were found at six sites for Promotion of Parent Involvement, while no net negative effects were observed. This was by far the most generalized effect unearthed by the evaluation. In sum, these findings indeed suggest that PDC may have been more effective for parents in general than for the small number of parents included in the longitudinal evaluation. However, it should be pointed out that "parents in general" were not predominantly parents of Head Start children, the focus of the intervention.

Regarding triangulation of findings across institutional, parent, and teacher/classroom outcomes, perfect consistency was observed in only one case--Site 8--where there were no effects of consequence found at any level of the design. Site 8 was also the only place where findings for parent outcomes agreed with findings for the related teacher/classroom outcome domain. However, partial confirmation of positive expectations based on institutional outcomes was achieved at all six sites for which such expectations were expressed--Sites 1, 2, 5, 6, 7, and 9. Given apparent differences between the parent samples analyzed here and in Chapter IV as well as the low power of our statistical tests, this latter analysis of the fit between predictions and outcomes may be the most appropriate. If so, it provides fairly strong evidence of PDC-favoring effects on parents at those six sites where institutional provisions for parent involvement/training were greater in PDC than local Comparison schools. PDC-favoring net effects for at least one of the three parent-related outcome domains were also found at three sites--Sites 3, 4, and 10--where no group differences

were predicted. Although these effects are not supported by institutional outcomes, and are contradicted by negative effects in one of two parent outcome domains at Site 3, they fit the overall pattern of generalized positive effect. Only Site 8, where the basic PDC program was diffused throughout the school district, shows no sign of group differences in parent-related outcomes.

Before moving on, it seems worth pausing to consider the magnitude of PDC's impact on the time that parents actually spent in the classroom. Mintues of Parent Time in Class was our only measure of parent involvement that did not rely upon parent or teacher self-report. The estimate was based on observations made over the course of one entire, arbitrarily chosen day in each classroom. The *possible (?)* PDC-favoring effect from the comparison of aggregate PDC and Comparison samples was based on underlying observed means of 8.2 minutes for PDC classrooms and 3.3 minutes for Comparison classrooms. From 1979 through 1981 parents were observed in only 8.5% (18/211) of Comparison classrooms and 17% (18/105) PDC classrooms. The average length of parent visits to PDC classrooms was 48 minutes; the average visit to Comparison classrooms lasted 39 minutes. Though surely some classrooms were almost never visited by parents while others had parent volunteers on hand nearly every day, one can think of the hypothetical average PDC classroom as having a parent visitor every 6 days, staying an average of 48 minutes, while the hypothetical average Comparison classroom was visited by a parent every 12 days, staying an average of 39 minutes. Whether the observed average level of parent involvement in PDC classrooms should be considered "high" or "sufficient" seems a matter of judgment.

PDC-Encouraged Instructional Content. Drawing upon its Head Start experience and guided by its concept of developmental continuity, ACYF made a fundamental commitment to promoting multicultural education, health and nutrition education, and the incorporation of community resources (materials and people, including parents) into the educational process. Net positive effects of consequence for the domain as a whole were found in three sites--Sites 2, 6, 8--and over all sites. Comparison-favoring effects were found at Sites 1 and 7. These findings do not strongly confirm ACYF's hopes; however, limited institutional data suggested that a generalized PDC-favoring effect would not necessarily be expected.

A primary mechanism for increasing the emphasis given these educational objectives in PDC classrooms was teacher training. The PDC Guidelines specifically required training in bilingual/bicultural/multicultural education (as appropriate) and health/nutrition/safety education (see Chapter III). *Post hoc* analyses of the relevant subcomponents (Tables 14, 15, 20, and 21 in Appendix D) indicated that appreciably more training in multicultural education occurred in PDC at Sites 5 and 7, while more training in health/nutrition education occurred in PDC at Sites 4 and 7. Amount of training represents the actual amounts of training reported by PDC and Comparison teachers who were interviewed. Findings reported in Tables V-5 and V-6 do not confirm expectations based on amounts of training provided. No effects of consequence were found for Sites 4 and 5, and net negative effects were found for Multicultural Instruction (Table IV-6) and Health/Nutrition Instruction (Table IV-4) in Site 7. We have no plausible explanation for these findings.

Domains Representing Generally Desirable Outcomes

Classroom Environment. The variables in this domain measure the least tangible aspects of teacher-child behavior and physical environment considered in the evaluation. Though Stimulating & Attractive Environment and Supportive & Enthusiastic Climate measure ostensibly different things, scores correlated moderately (.41) and together they seemed to characterize the relative desirability of classroom settings from the perspectives of observers who were themselves teachers. Interobserver/rater agreement was high during the training period.

Though no specific expectations were generated by institutional outcomes, we did not anticipate finding Comparison-favoring effects. Insofar as we know, the material resources available to PDC schools and teachers were at least equivalent to resources available in Comparison schools. And PDC placed special emphasis on creating school and classroom settings that would reduce "environmental discontinuity" and support the social-emotional, as well as academic, development of the child. Moreover, schools seeking to increase the involvement of parents and other members of the community might be expected to pay special attention to the impression their classrooms would make upon visitors.

Curiously, consequential net negative effects were found at two sites (2 and 3); a net positive effect, at only one site (8). No effect of consequence was found over all sites. It is somewhat disturbing that both Comparison-favoring net effects were due, at least in part, to Comparison classrooms having been rated higher on Supportive & Enthusiastic Climate, the more compelling of the two variables in this domain. We have no explanation of the findings.

Educational Management. Though one may disagree with the way educational management was operationalized, or the quality of measurement, it is unlikely that one would dismiss the objectives of measurement in this domain. And though PDC Guidelines did not prescribe any of the specific outcomes measured, positive effects would seem desirable and compatible with PDC's larger goals, while negative effects would appear to cast doubt on the value and effectiveness of the intervention.

Unfortunately, consequential net negative effects were found at two sites (1 and 3) and over all sites, with only one site (6) howing a PDC-favoring effect. The overall Comparison-favoring effect appears to have been caused by rather strong negative patterns at Sites 1 and 3 in conjunction with negative trends in Sites 2 and 4. There were no negative findings at any other site (5-10; Table V-6).

These findings were sufficiently disconcerting that considerable effort was made to find an explanation. One possibility we considered was that open-plan classrooms and small-group seating arrangements might be associated with some aspects of management difficulty and occur more frequently in PDC than Comparison schools. Anecdotal data picked up in conversations with observers over three years of data collection suggested that some found less conventional classrooms chaotic and difficult to observe. Thus, if PDC contained more such classrooms it was possible that management would be more difficult and less successful, either in reality or in the eyes of the observers. In fact, the PDC sample contained significantly more open-plan classrooms and classrooms utilizing small-group seating arrangements (whether or not they were open-plan). However, when these two features of classroom structure and organization were correlated with Educational Management outcome variables, they were found to predict more successful, rather than less successful management. Correlations were consistently positive, albeit weak (<.25) in the total sample and in the aggregate PDC and Comparison samples, contradicting our commonsense.

We then turned to available data on teacher background. As discussed earlier in this chapter, we had decided not to include teacher background and demographic characteristics as covariates in Design 1 and 2 analyses for two primary reasons: most did not appear to be related to teacher/classroom outcomes, and those that were related to outcomes were implicated by anecdote as dimensions along which self-selection into and out of the program had occurred. The best predictor of outcomes was "previous teaching experience," particularly previous experience teaching at the same grade level, and the average experience of PDC teachers was found to be significantly lower at several sites and over all sites.

In order to determine whether differences in the teaching experience of PDC and Comparison teachers might explain observed differences in Educational Management, we reran Design 1 and 2 analyses for these outcomes including "number of years previous teaching experience at same grade level" as part of the basic covariate set. Net negative effects of the same magnitude persisted at Sites 1 and 3, and the consequential net positive effect at Site 6 was unaltered. However, a second net positive effect of some magnitude (+.67) appeared at Site 8, which formerly registered a net negative effect of -.25 based on *probable* findings and a zero effect when *possible (?)* findings were added (Table V-3). Moreover, the net negative effect of -.33 found for the aggregate samples in the original analysis disappeared. Thus, while inclusion of teaching experience as a covariate did not entirely eliminate negative findings, it shifted the balance of site-level effects and altered the overall trend.

It would seem, then, that PDC difficulties with educational management can in part be traced to the relative inexperience of PDC teachers at some sites--e.g., an average 2.5 years prior experience at grade level for PDC teachers in Site 8 versus 6 years prior experience at grade level for local Comparison teachers. But inexperience alone does not account for the net

negative effect observed in Site 1, where PDC teachers were also significantly less experienced than Comparison teachers. And inexperience played no part in the Comparison-favoring effect at Site 3, where PDC and Comparison teachers were equally highly experienced. No other plausible explanations could be found.

It is difficult to know how best to interpret the findings from this re-analysis of the Educational Management domain. Should one view differences in teaching experience as a source of bias in outcome measures--i.e., as a pre-existing difference between the PDC and Comparison groups that "explains away" what appeared to be a generalized negative effect leaving us with only two aberrant cases to ponder? Or should one view group differences in previous teaching experience as the product of an intervention that created selection pressures which ultimately altered teacher demographics in a direction potentially unfavorable to the program's larger goals? Surely the answer is not so simple as either question implies. We regret that available information does not allow us to shed more complete and certain light on the situation.

Learning Time. In discussing effects in this domain we use the terms "negative" and "positive" with some trepidation. In defining the nature and purposes of PDC, ACYF said nothing specifically about increasing allocations of teaching-learning time to specific academic subjects, much less about how much time should be allotted to different content areas of the curriculum.

Yet, one of PDC's stated objectives was to support the development of children's basic academic skills; therefore, it would not seem unreasonable to examine instructional process variables that have been demonstrated to bear upon academic achievement as measured, however narrowly, by conventional achievement tests. Our rationale for calling differences in learning time "negative" or "positive" (favoring/disfavoring) is not simply that negative represents less time allocated and positive more. Rather, we assume that given the range of natural variation in PDC communities, higher scores on Learning Time outcome measures generally reflected more effective teaching practice. While recognizing that increased allocations of time to conventional math instruction, for example, must eventually reach a point of diminishing returns as judged by children's levels of attention and rates of learning, we seriously doubt that many classrooms in the evaluation sample were approaching such a point. Thus, when PDC children at a particular site were observed to spend less time, or less engaged time, in specific academic activities than Comparison children in the same community, we assumed that PDC children were probably not spending enough time to keep up with other children in their communities, not to mention national norms, over the long haul. And this we considered a negative effect. It is important to remember in this respect that differences in site-level group means had to be quite large (over ½ a standard deviation

in many cases) to be detected by our design. (At Site 3, for example, the probable net negative effect for Minutes/Child Engaged in Reading with High Attention was based on a difference in observed means of 30 minutes/day/child--approximately ½ a standard deviation for that sample and 1/8 of the class day.)

The pattern of effects for the Learning Time domain resembled the pattern for Educational Management, but was more generally negative with Comparison-favoring net effects at four sites (2, 3, 4, 8) and over all. A PDC-favoring effect was found at one site (5). Findings were consistent across the two domains only at Site 3 where a sizable net negative effect was found in both instances.

We were surprised that there was not greater consistency between findings for Learning Time and Educational Management. However, correlational data indicated that the relationships between outcomes in the two domains, though consistent with our logical analysis, were not particularly strong. Two management variables--Percent of Teacher Time Spent Teaching and Percent of Child Time Engaged in Any Educational Activity--correlated weakly to moderately, and positively, with all Learning Time outcomes. Three--Maintenance of an Orderly Classroom Process, Level of Teacher and Child Involvement in Learning, and Percent of Child Time Engaged in Learning with High Attention--correlated moderately to strongly with Minutes/Child Engaged in Math and Engaged in Reading/Language Arts with High Attention. The remaining management variable--Management of Information for Individualization--was essentially unrelated to Learning Time outcomes. Perhaps part of the explanation for lack of fit between outcomes for the two domains was that teachers, even those who were excellent managers, did not control the amount or blocking of time they had to manage, but worked within parameters established by the larger system. Thus, a teacher who succeeded in keeping children engaged at high levels of attention (Percent of Child Time Engaged in Learning with High Attention) during math instruction might not achieve as many minutes per child with high attention (Minutes/Child in Math with High Attention) as a less effective teacher simply because the block of time available for instruction was limited by factors beyond her control.

As with Comparison-favoring findings for Educational Management, we wanted very much to understand the origins of apparent negative effects in this domain. Toward that end we evaluated the same classroom structure and organization variables, and the same teacher background characteristics as potential explanatory factors for outcomes in this domain. In addition we investigated the possibility that PDC/Comparison differences in Minutes/Child Available for Learning (the length of the class, but not the school day) might account for differences in the amounts of time allocated to different learning activities. Only "previous teaching experience" and "length of class day" were found to correlate significantly and fairly consistently with Learning Time outcomes. When we added these variables to our analytic design as covariates, only one net effect was changed. In

Site 2 a moderate net negative effect dropped below our cutoff for consequentiality; addition of the covariate "previous teaching experience" was responsible. As in the previous section, we seriously question whether or not a mean difference between PDC and Comparison teachers in teaching experience should be considered a biasing factor rather than a program effect accomplished by selection.

One additional possibility was also entertained--that local PDC programs systematically broadened their curricula to address subjects not generally dealt with in local elementary schools; to devote more time to projects that were not strictly academic; or to promote activities (family-style meals, discussions, games, etc.) intended to foster social development and general communication skills. If this were the case, teachers might well have cut back on time allocated to math, reading/language arts, science, and formal social studies instruction to make room for other activities that were considered fundamental to the PDC program, which afterall was concerned with developing children's general social competence. Although available data do not permit us to dig further by any means but unsystematic anecdote, this explanation is certainly plausible.

Finally, it should be pointed out that our assumption that negative effects in this domain would have unfavorable implications for children's achievement was just that, an assumption. When evaluating child outcomes in the next chapter, we shall consider the fit between the findings reported here and children's academic achievement in math and reading through third grade.

VI

CHILD OUTCOMES

The ultimate goal of PDC was to enhance children's social competence--their everyday effectiveness in dealing with the environments of home, community, and school. In this chapter, we examin- measured child outcomes for evidence of such program effects, addressing our fourth research question:

> Q4: Did PDC children exhibit greater social competence than would have been expected had they not participated in the program?

Given the pattern of findings for institutions (Chapter III), parents (Chapter IV), and teacher/classrooms (Chapter V), there was no reason to expect generalized positive or negative effects across sites or outcome domains within sites. Expectations were particularly difficult to generate for child outcomes because of the lack of either tight logical-theoretical or well-established empirical fit between the specific outcomes measured for children and most of the outcomes measured at the level of parents and teachers/classrooms. Thus, we embarked upon the evaluation of child outcomes with an open mind, expecting very little but prepared to entertain strong patterns of effects if they occurred. And if systematic program effects on children were indicated, it would then make sense to address our fifth research question:

> Q5: Which program effects on teachers and parents account for program impacts on children?

In addressing this question, we would, of course, be limited to consideration of those parent and teacher impacts that were measured.

This chapter describes first the measurement of child outcomes, then the data analytic strategy used to estimate program effects. In the third section, findings are presented and interpreted.

MEASUREMENT OF CHILD OUTCOMES

The primary objective of instrument selection and development was to achieve reliable and valid measurement of children's *social competence*. Other criteria considered in forming the child measurement battery included (1) suitability of instruments for administration by paraprofessionals, (2) the time required for administering particular instruments and the entire battery, (3) the age appropriateness of measures and the possibility of continuous measurement from Head Start through third grade, (4) the availability of Spanish language versions or the ease with which items could be translated, and (5) the history of instrument use in other national evaluations. Lacking an adequate operational definition of social competence and constrained by time and available resources, it was not possible to

develop a battery that convincingly measured the many facets of children's social competence. However, it is hoped that the PDC child battery represents a positive step in that direction by comparison with previous large-scale evaluations.

Baseline Measures

PDC children officially entered the program when they enrolled in Head Start in the fall of 1976. Both PDC and Comparison children attended Head Start for one year prior to entering kindergarten, and in two sites, they attended the same centralized Head Start centers. All children were tested and observed shortly after beginning Head Start in fall 1976 and again at the end of the Head Start year in spring 1977. Fall measurements were intended to provide a "pretreatment" characterization of children; spring 1977 measurements, to document possible program effects during the Head Start year.

Interim Report VII, Volume 3 (Granville, Love, & Morris, 1977) closely examined child outcomes as of spring 1977 in relation to treatments experienced during the Head Start year. The authors of that report concluded that there was no indication of systematic program effects at the Head Start level and no indication that PDC and Comparison children had systematically different Head Start experiences within sites.

Given these findings, a decision was made to treat all child measures from the Head Start year as *baseline* measures of children in the evaluation cohort. Though we are fairly confident that neither fall nor spring performance was influenced by the PDC intervention, we avoid calling the aggregate Head Start measures "pretreatment measures" because formally they are not. The major reason for deciding to aggregate all Head Start child measures into baseline estimates of child characteristics was to increase the reliability of variables used to evaluate the initial comparability of child samples and to control for noncomparability in analyses of covariance. Fall and spring scores for each instrument were standardized separately for the English- and Spanish-dominant samples. Fall scores were then averaged with corresponding spring scores for each child. Children missing data for a particular measure at one or the other testpoint were assigned the single available score as their "average." This procedure permitted the inclusion of children missing fall 1976 data, but having spring 1977 data, in analyses requiring baseline measures, an important consideration given small site-level sample sizes. Correlations between Head Start measures and later outcomes were consistently higher when average scores, rather than fall 1976 or spring 1977 scores, were used, suggesting that averaging fall and spring scores did increase the reliability of early assessments of child characteristics.

The Head Start measurement battery and its psychometric characteristics are described at length in Interim Report IV, Volume 1 (Granville, McNeil, Meece, Wacker, Morris, Shelly, & Love, 1976). Here, we only provide brief descriptions of each instrument considered in the longitudinal evaluation.

Outcome Measures

Measurements made from spring of 1978 (kindergarten year) through spring of 1981 (third grade year) are considered *outcome measures.*[1] The battery gradually changed as children moved into and through the elementary grades to reflect the changing definition of social competence as defined by public elementary schools.

Data Collection Design

The data collection design for child measures considered in the longitudinal evaluation is illustrated in Table IV-1. The abbreviated instrument names used in this table will be used throughout and will be explained in subsequent paragraphs. Only four measures were obtained at all testpoints from fall Head Start through spring third grade (1981)--BSM, PIPS, and POCL 1 and 2.

With the notable exception of Block Design at spring of Head Start (1977) and the entire column for spring of kindergarten (1978) cell N's hover at or above 300 (total longitudinal sample = 326). The Block Design test was only administered at four sites in spring 1977 in order to obtain data for estimating score stability and for evaluating the psychometric structure of the battery. The median cell N of 277 in the spring 1978 (kindergarten) column of Table IV-1 is the product of having excluded data for 25 children tested in Spanish at Site 8. The only children remaining in the longitudinal sample who were judged to be dominant in Spanish at entry to Head Start resided in Site 8. All of these children were tested in Spanish during Head Start, and approximately one half of them were also tested in Spanish at the end of kindergarten, based on judgments by teachers that they were not sufficiently bilingual to be tested in English. Rather than include their scores from the Spanish version of the child battery, which did not appear to be equivalent to the English version, a decision was made to exclude these children from overall analyses for spring 1978 but to include them again in spring 1979 when scores from English language versions of all tests were available for all children. Separate, and quite exploratory, analyses of bilingual program effects at Site 8 will be reported later in this chapter.

---

[1]Throughout this chapter years and grade levels are used interchangeably. However, the reader should be aware that some children in the evaluation sample were not on grade. One percent of the longitudinal analytic sample (n=326) had been retained one year as of 1978; 3%, as of 1979; 10%, as of 1980; and 17%, as of 1981 (the nominal third-grade year). No children remaining in the longitudinal analytic sample had been retained for more than one year as of spring 1981.

Table IV-1

Child Measures Considered in the Longitudinal Evaluation and Available Data at Each Testpoint for the Total Longitudinal Sample (N=326)

	BASELINE		OUTCOMES			
	Head Start Fall 1976 N	Spr. 1977 N	Kdg. Spr. 1978 N*	G1 Spr. 1979 N	G2 Spr. 1980 N	G3 Spr. 1981 N
Specific Academic Achievement						
PIAT Math	-	-	279	302	298	316
PIAT Reading	-	-	272	293	-	-
MAT Reading	-	-	-	-	311	314
General Academic Skill/Aptitude						
Block Design	325	102	-	-	-	-
BSM	321	313	277	310	313	326
Verbal Fluency	324	314	277	310	-	-
Verbal Memory 1	324	315	277	-	-	-
Verbal Memory 2	319	312	277	-	-	-
Draw-a-Child	323	315	277	-	-	-
Learning Attitude/Style						
POCL 1: Task Orientation	326	297	255	291	288	325
CI 2: Interest in Reading	-	-	-	305	305	326
CRS 3: Learning Orientation	-	-	-	304	309	315
Attitude toward Teacher/School						
CI 1: Attitude toward School	-	-	-	308	312	326
PI: Attitude toward School	-	-	-	266	279	279
School Attendance	-	-	-	292	278	260
Social Development/Adjustment						
PIPS	324	314	277	309	313	326
POCL 2: Sociability	326	297	255	291	288	310
CRS 1: Independence	-	-	283	304	309	313
CRS 2: Social Adjustment	-	-	283	304	309	315

*Excludes 25 children tested in Spanish at Site 8. Their inclusion would push cell N's over 300 in all but two cases.

Data collection procedures (hiring, training, monitoring) are described in Appendix E. Score distributions (histograms) by outcome variable by year by group, intercorrelations of outcome variables at each data collection point and across years, and estimates of the internal consistency of outcome measures (when appropriate) across years are presented in Appendix H.

Instrument Descriptions

Instrument descriptions are grouped in the same way that measures were conceptually grouped in Table IV-1.

Specific Academic Achievement

Instruments in this group were first administered in spring 1978 as children completed kindergarten. The instruments selected were intended to measure how well children could cope with specific academic tasks in the areas of numeracy and literacy. Development of skills in both of these areas is given high priority in virtually all public elementary school programs beginning in the kindergarten year. Moreover, mastery of these skills is an essential feature of social competence in the school environment. (This is true whether or not one agrees with the high priority that most schools place upon the acquisition of such skills in the early childhood period.)

PIAT Math. The mathematics subtest of the Peabody Individual Achievement Test (PIAT: Dunn & Markwardt, 1970) was administered from kindergarten through third grade. Published instructions for presentation and scoring were followed to the letter. The test was administered to children individually. The examiner read questions aloud, and children responded by pointing to one of four answers displayed on a printed plate in the test book; thus, reading skill was not confounded with mathematics skill. Not all items were administered to each child; rather, the examiner determined a basal performance level for each child, then presented items in a sequence of increasing difficulty until the child reached his/her performance ceiling. Items progressed in difficulty from non-numerical visual discrimination tasks to computational story problems. Raw scores (number of items up to basal + number of items correct from basal to ceiling) were analyzed.

The publisher reported one-month test-retest correlations of .52, .83, and .68 at kindergarten, first grade, and third grade, respectively. One-year test-retest correlations for the PDC evaluation sample were comparable--.59, .70, and .72 for kindergarten-first, first-second, and second-third, respectively.

PIAT Reading. PIAT Reading refers to the Reading Recognition subtest of the Peabody Individual Achievement Test (PIAT: Dunn & Markwardt, 1970). This test was administered in spring kindergarten and spring first grade. It was selected to provide a measure of early reading and reading-related skills. As with PIAT Math, published administration and scoring procedures were followed. The test was administered to children individually, and having established a basal performance level, the examiner only

administered items up to the child's performance ceiling. Items progressed in difficulty from nonlinguistic visual discrimination tasks to reading isolated words aloud. Raw scores (number of items up to basal + number of items correct from basal to ceiling) were analyzed.

The publisher reported one-month test-retest correlations of .81 at kindergarten and .89 at first grade. The one-year (kindergarten-first) test-retest correlation for the PDC evaluation sample was .71.

MAT Reading. MAT Reading refers to the Reading Survey Test of the Metropolitan Achievement Tests (MAT: Prescott, Balow, Hogan, & Farr, 1978a and 1978b). It was administered in spring second grade and spring third grade. The MAT Reading Survey was selected to measure reading comprehension skills not tapped by the PIAT Reading Recognition test. Two levels of the MAT Reading Survey were employed: the Primer at second grade; the Primary 2 at third grade. Since the Primer was developed and normed for kindergarten, second graders in the evaluation sample were tested two years below level in spring 1980. The Primary 2 was developed and normed for second grade; thus, third graders in the evaluation sample were tested one year below level in spring 1981. Whenever possible, the MAT Reading was administered to groups of 4 to 7 children belonging to the evaluation sample and attending the same school. Published administration and scoring procedures were followed. Raw scores (total number of items correct) were analyzed.

The publisher reported reliability coefficients (K-R 20) of .85 and .95 for the Primer and Primary 2 norm groups, respectively. Internal consistency[1] for the PDC sample was .88 and .94 at second and third grades, respectively. The one-year (second-third) test-retest correlation for the PDC sample was .71, and PIAT Reading scores from spring first grade correlated .56 with MAT Reading scores from spring second grade.

General Academic Skill/Aptitude

Instruments in this category were selected to measure a broad range of cognitive-linguistic skills thought to have general application in academic learning and in problem-solving both in and out of school. Of the six tests considered here, only one (BSM) appeared to provide a direct measure of social competence--"everyday effectiveness"; the others, in varying degrees, presented children with tasks not generally encountered and/or not similarly evaluated in everyday life. All of the instruments in this set were administered in either Spanish or English during the Head Start year, depending upon the child's language dominance. As previously noted, a small number of children were administered Spanish versions of some or all tests in the spring of their kindergarten year as well.

Block Design. The Block Design test used in PDC was adapted (with permission) from the Block Design Subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI: Wechsler, 1967). The content,

[1]Throughout, we have estimated internal consistency using Cronbach's alpha.

presentation, and scoring of the PDC version are described in Appendix H. The Block Design test was administered at all sites in fall 1976, when children entered Head Start, and at four sites in spring 1977 (for purposes of psychometric studies of the child battery). The task required children to construct designs using flat, colored blocks (taken directly from WPPSI test kits), copying either a model constructed by the examiner or a picture of a design. The Block Design test was intended to measure problem-solving abilities, flexibility of response style, and visual-motor organization--skills that might have generalized to diverse "real world" tasks. Raw scores (number of designs correct) were analyzed.

Internal consistency of Block Design scores for the evaluation sample (English and Spanish samples analyzed separately) ranged from a low of .75 at fall of Head Start to a high of .81 in the spring of the Head Start year. The test-retest correlation from fall to spring of Head Start (for the English-dominant children in the four sites where data were collected) was .61.

BSM. The Bilingual Syntax Measure (BSM--Burt, Dulay, & Hernandez Ch., 1975 and 1978) was administered at all sites at all testpoints. An attempt was made to administer the English version--BSM English--to both English- and Spanish-dominant children beginning in the fall of their Head Start year. All Spanish-dominant children were also administered the Spanish version of the test (BSM Spanish). Because very few valid BSM English performances were elicited from Spanish-dominant children during the Head Start year, only their BSM Spanish scores were included in the baseline data set. Only BSM English scores, for both Spanish- and English-dominant children, were included in the main outcome analyses reported here.

The BSM was intended by its authors to measure children's "relative proficiency with respect to basic syntactic structures" in either English or Spanish. The BSM elicited children's responses to questions asked by the examiner in a structured dialogue centered around cartoon pictures. Children's responses were evaluated according to the appropriateness of their content, syntactic structure, and language (children were expected to respond in the language--Spanish or English--used by the tester). According to the authors, scoring was supposed to take the child's dialect into account by not marking nonstandard syntactical structures as incorrect when they represented acceptable usage among adult members of the child's language community. Under such scoring conditions, the BSM would have represented a measure of children's functional oral language competence in the dialects of their homes and communities. However, the actual scoring procedure used in PDC did not systematically allow for departures from the "standard" dialects of the scorers and, consequently, resulted in a measure that confounded dialect and code-switching behavior with oral language proficiency and mastery of syntax. Thus, BSM scores would appear to measure functional oral language competence in the Spanish or English dialects preferred by schools (insofar as our well-educated scorers represented these dialects). Moreover, since no effort was made to encourage children to respond in standard English or Spanish dialects (in fact, an attempt was made to match children with paraprofessional testers from their own ethnic/linguistic communities to put them at ease), it is not clear how well the BSM measured children's mastery of standard dialect.

Internal consistency for the Spanish-dominant sample was .86 in the fall of Head Start and .70 in spring; the fall-spring correlation was .59. For the English-dominant sample, internal consistency was .83 in both fall and spring of the Head Start year, and the fall-spring correlation was .71. Internal consistency of BSM English for the total sample ranged from .68 to .75 from spring kindergarten through spring third grade. One-year test-retest correlations from spring Head Start through spring third grade ranged from .56 to .68.

Verbal Fluency. Verbal Fluency was derived from a subtest of same name in the McCarthy Scales of Children's Abilities (MSCA: McCarthy, 1972) and was used in the PDC evaluation with permission from the publisher. It was administered from fall Head Start through spring third grade. The instrument was intended by the author to measure children's classification skills. It seems likely that the instrument measured a combination of general knowledge, information retrieval and classification skills, oral language skill, and the child's "presence of mind." Children were asked to name as many different members of a noun class designated by the examiner as they could within 20 seconds. Four categories were used: "animals," "things to eat," "people's names," and "toys." The first two categories were taken directly from the MSCA subtest; the latter two categories were substituted by the evaluator in an effort to increase the instrument's cultural relevance for children in the evaluation sample. The evaluator also modified, in order to simplify, MSCA scoring procedures. Presentation and scoring of the PDC version are described in Appendix H. Raw scores (total number of different and relevant nouns named) were analyzed.

The test's internal consistency for the Spanish-dominant sample was .81 in the fall and .76 in the spring of the Head Start year; the correlation between fall and spring scores was .52. For the English-dominant sample, internal consistency was .76 at both Head Start testpoints, and the fall to spring correlation was .49. Internal consistency for the total sample dropped at spring kindergarten to .68, then increased to .76 again by the end of first grade. One-year test-retest correlations were .52 and .41 for Head Start-kindergarten and kindergarten-first, respectively.

Verbal Memory 1. Like Verbal Fluency, Verbal Memory 1 was derived from an MSCA subtest of the same name (McCarthy, 1972--by permission). It was administered from fall Head Start through spring kindergarten. The instrument was intended to measure one facet of children's short-term verbal memory. Children were asked to repeat sequences of unrelated words that had been read by the examiner. Of the seven word sequences used, three were taken directly from the MSCA subtest, and four were created by the evaluator, replacing MSCA items requiring children to repeat sentences rather than disparate words. Scoring also differed slightly from published procedures. Content, presentation, and scoring of the PDC version are described in Appendix H. Raw scores (number of words correctly recalled in order) were analyzed.

For the Spanish-dominant sample internal consistency was .84 at fall Head Start and .85 in the spring; the fall-spring correlation was .69. For the English-dominant sample internal consistency was .82 in both fall and spring of the Head Start year; the fall-spring correlation was .51. For the total sample at spring kindergarten, internal consistency was .73, and the correlation of kindergarten with spring Head Start scores was .53.

Verbal Memory 2. This test was, in fact, Verbal Memory, Part II of the MSCA (McCarthy, 1972). It was administered only through kindergarten. Like Verbal Memory 1, the test was intended to measure short-term memory, but memory of connected discourse in story form rather than unrelated words. Raw scores (number of pieces of information correctly recalled not necessarily in order) were analyzed.

Internal consistency for the Spanish-dominant sample was .84 in the fall and .78 in the spring of Head Start; the fall-spring correlation was .39. Internal consistency for the English-dominant sample was .82 in the fall and .81 in the spring of Head Start; the fall-spring correlation was .47. Internal consistency for the total sample at spring kindergarten was .75; the one-year (Head Start-kindergarten) test-retest correlation was .39.

Draw-a-Child. This test was also adapted from a MSCA subtest of the same name (McCarthy, 1972) and was administered only through kindergarten. Task presentation followed MSCA procedures: the child was asked to draw a picture of a child (boys drew boys; girls, girls). In the MSCA scoring system, credit is given both for inclusion and for definition of ten body parts. In the PDC version, scoring was simplified to note only presence/absence of each part. Such tasks were originally used as measures of "general intelligence"; however, McCarthy viewed Draw-a-Child as an index of "perceptual performance," which might be expected to generalize to performance in more ordinary social settings. Raw scores (total number of required body parts represented) were analyzed.

Though the appropriateness of estimating instrument reliability using estimates of internal consistency might be questioned in the case of this test (since the drawing task can be viewed as a single item), coefficients were computed. For the Spanish-dominant sample, internal consistency was .78 in the fall and .69 in the spring of Head Start; the fall-spring correlation was .39. Internal consistency for the English-dominant sample was .83 at fall and .81 at spring Head Start; the correlation from fall to spring was .54. Internal consistency for the total sample at spring kindergarten dropped to .68, and the correlation from spring Head Start to spring kindergarten was .48.

Learning Attitude/Style

In contrast to most measures in the preceding category, the instruments considered here were intended to provide direct estimates of children's functional competence, attitudes, and motivations as learners/problem-solvers in and out of school.

POCL 1: Task Orientation. The High/Scope Pupil Observation Checklist (POCL) used in the PDC evaluation was adapted from an instrument of the same name developed and used by the High/Scope Educational Research Foundation in previous evaluations (e.g., Love, Nauta, Coelen, Grogan, McNeil, Rubin, Shelly, & Stein, 1976). The POCL 1 was completed for each child at all testpoints by testers, after they had administered all tests in the battery. Assessments were based upon observations of children in varied test and interview settings. The POCL required testers to rate children on a set of seven-point scales defined at their extremes by bipolar adjectives. (The checklist is reproduced in Appendix H.) POCL 1 was the sum of ratings on eight scales that hung together empirically (based on factor analyses) and appeared (commonsensically) to measure a child's approach to and manner of engaging in tasks of the sort presented by tests and classroom instruction.

Internal consistency from fall Head Start through spring of third grade never dropped below .92. One-year test-retest correlations ranged from .22 to .44 in no particular relation to grade level.

CI 2: Interest in Reading. Scale 2 of the PDC Child Interview (CI) was derived from procedures developed for an evaluation of High/Scope-sponsored Follow Through projects in 1976-77 (Kittel, Tamor, Smith, & Bond, 1977). The part of the interview concerned with reading behavior was conducted as an informal conversation between interviewer and child, in which the interviewer asked a number of standard questions to stimulate and guide the conversation. After the necessary information had been elicited from the child, the interviewer completed a number of ratings, each on a scale from 1 to 5. Three of these ratings were then averaged to produce the CI 2. These ratings estimate the amount of reading a child engages in spontaneously and the child's perception of reading as something more than an instructional activity. (The procedure is described further in Appendix H.) The Child Interview was conducted at three testpoints--first through third grades.

Internal consistency was .94 at all grade levels; one-year test-retest correlations were .35 and .41 for first-second and second-third, respectively.

CRS 3: Learning Orientation. The Learning Orientation variable was derived from the PDC Child Rating Scale (CRS), which was completed by teachers of children in the evaluation sample from first through third grades. CRS items were presented as descriptions of a child's behavior. Teachers were asked to judge how characteristic each description was of the child in question, indicating how frequently the child behaved in the manner described in comparison with other children of the same age and background. The Learning Orientation measure was an average of seven ratings from the CRS. Items were adapted from various sources (Vinter, Sarri, Vorwaller, & Schafer, 1966; Bloom, 1976; Weikart, Bond, & McNeil, 1978). The specific items used and their interrelationships are described in Appendix H. The Learning Orientation measure was thought to provide an index of a child's general orientation (attitude, motivation, approach) toward learning and probably an index of the teacher's assessment of a child's "educability" within the school context.

Internal consistency was .93 at first and second grades and .92 at third grade; one-year test-retest correlations were .54 and .48 for first-second and second-third, respectively.

Attitude toward Teacher/School

Measures of children's attitudes toward teachers and schools were included in the outcome battery because positive attitudes were thought to be (1) prerequisite to effective functioning in the school environment and (2) the products of school and home environments that enhanced general social competence.

CI 1: Attitude toward School. Like Interest in Reading, this measure was derived from the PDC Child Interview (CI) conducted in first through third grades. The elicitation procedure was adapted (with permission) from the "attitude toward school" scale of the Purdue Social Attitude Scales (Cicirelli, 1969). The PDC version incorporated items 2, 6, 9, 12, 16, 19, 26, and 28 of the published scales. The child responded to picture stories in which the stick figure representing the main character was given the child's name. A figure representing the child's teacher also appeared in all but one of the stories. The interviewer started each story, identifying the characters and creating a potentially dramatic situation. The child was then asked to finish the story by selecting from among five faces, with expressions ranging from very happy to very sad, the one face that best represented the reaction of a particular character--either the child's reaction to a school situation or the reaction of the teacher or principal to the child.

Internal consistency was .54 at first grade and .47 at second and third grades; one-year test-retest correlations were .17 and .26 for first-second and second-third, respectively.

PI: Attitude toward School. Another measure of the child's attitude toward school was obtained in the PDC Parent Interview (PI) conducted from first through third grades. The PI attitude measure was based on a single item in which parents were asked to indicate "how true" (1=definitely true; 5=not at all true) the following statement was: "(child's name) loves school and enjoys being there." No estimate of internal consistency could be made; however, correlations over one-year intervals were .55 and .38 for first-second and second-third, respectively--suggesting moderate stability in parents' assessments of their children's attitudes at least.

School Attendance. School attendance data were collected from school records from first through third grades. Though a child's rate of attendance was surely influenced by factors other than the child's attitude toward school, it was thought that differences in group means would likely reflect differences in children's attitudes, perhaps confounded with parental attitudes toward school. The School Attendance variable was constructed as a ratio of days attended divided by the total number of days that the child's school was in session in a given year.

Social Development/Adjustment

Instruments in this category were intended to provide measures of children's functional competence in social interactions with peers and adults.

PIPS. The Preschool Interpersonal Problem-Solving test (PIPS: Shure & Spivack, 1974) was adapted for use in PDC with permission of the authors and was administered from fall Head Start through spring of third grade. The PDC version included six items from the "problem between peers" section of the original PIPS and added one new item. The examiner used pictures of two children and various toys to create dramatic situations in which one child was playing with a toy, and the other wanted to play with it. The child being interviewed was asked to say how the child who wanted to play with the toy might get a chance to do so. Children were encouraged to generate different solutions in response to seven different situations, each involving a different toy. If a child was unable to generate a new solution, the test was terminated at that item. Children's scores were the number of distinct solutions offered. Higher PIPS scores were thought to indicate that children had larger social problem-solving repertoires and would more likely find appropriate solutions to problems with peers. The actual scoring procedure differed slightly from that developed by Shure and Spivack (1974). Content, presentation, and scoring of the PDC version are described in Appendix H.

Since children were administered different numbers of items depending upon their abilities to generate distinct problem solutions, internal consistency could not be meaningfully estimated. One-year test-retest correlations ranged from .19 to .40 from fall Head Start through spring third grade.

POCL 2: Sociability. Like Task Orientation (above), Sociability was rated by testers using the High/Scope Pupil Observation Checklist (POCL) at all testpoints from Head Start through third grade. Ratings were made after all tests had been administered to the child. Sociability scores were formed by taking the sum of ratings for three POCL items (seven-point scales defined by bipolar adjectives--see Appendix H). POCL 2 scores were believed to indicate how relaxed and outgoing children were in the company of strange but friendly adults, a characteristic thought to have important implications for successful coping in more ordinary life settings.

Internal consistency ranged from .87 through .91 from Head Start through third grade; correlations over one-year intervals ranged from .30 to .59.

CRS 1: Independence. Independence was measured by two items from the PDC Child Rating Scale (CRS--see discussion of CRS 3 above and in Appendix H). The CRS 1 was completed by classroom teachers for children in the evaluation sample each spring from kindergarten through third grade. It is ostensibly a measure of a child's independence of other children, the extent to which the child can be said to "have a mind of his or her own."

Internal consistency ranged from .71 to .82 from kindergarten through third grade; spring-to-spring correlations ranged from .13 to .25.

CRS 2: Social Adjustment. Social Adjustment was measured by six items from the PDC Child Rating Scales (Appendix H). CRS 2 ratings were made by classroom teachers every spring from kindergarten through third grade. The CRS 2 appeared to measure children's social adjustment to the school context and probably provided an index of whether the teacher viewed a child's behavior to be problematic within the classroom setting.

Internal consistency ranged from .85 to .89; spring-to-spring correlations, from .48 to .52.

Overview of Child Measures

On the face of it, the child battery was strongly biased toward measurement of social competence in relation to the school environment. This was perhaps inevitable given existing instruments, the level of definition of the social competence construct, and the resources and time available for instrument development. In retrospect, this bias also seems to reflect the primary concern of persons involved with the PDC project at all levels--that Head Start children cope more effectively in public elementary schools.

Another observation that should be made of the child outcome battery is that there was considerable room for measurement bias--particularly on rating scales. In the course of analyzing child outcome data, we detected one type of measurement bias that fortunately did not prejudice site-level group comparisons, but indicated the potential for prejudicial bias. It appears that tester-raters were not always "calibrated" in the same way and, consequently, sometimes produced ratings that were much higher or lower for both PDC and Comparison groups at particular sites than would have been predicted given baseline-outcome relations in the total sample. For example, some testers' notions of what constituted "shyness" (POCL 2: Sociability) appeared to be quite different from the notions of others due presumably to cultural and personal differences among testers. Such "calibration" problems were likely to have afflicted teacher and parent raters as well; however, because teachers and parents were nested within program groups (actually within classrooms and families), the effects of any bias due to calibration problems among these raters could not be detected. It should be noted that none of the testers, teachers, or parents involved in measuring child outcomes was blind to either the purpose of the evaluation or the child's program affiliation. Though we have no reason to suspect that intended or unintended measurement bias systematically inflated or depressed scores for children in one group versus the other, the possibility cannot be ruled out.

Psychometric Properties of the Battery

The psychometric properties of the child outcome battery have not been well established. Estimates of reliability/generalizability were limited to measures of internal consistency (the degree to which individual responses to different items administered on the same occasion were intercorrelated) and stability (occasion-to-occasion correlations) usually over long periods of time (typically a year). Though internal consistency coefficients were generally encouraging, estimates of stability--i.e., the accuracy with which a particular score characterizes a child at other moments of measurement--were quite discouraging for many instruments. Of course, measurement-remeasurement intervals of one year are quite long, and a lot of "true score" change might occur during such a period. One can assume that the stability (generalizability across occasions) would have been higher for most measures over shorter intervals, but lacking actual data, one can only guess how much higher. Our guess, based on available data, psychometric theory, and experience, is that PIAT and MAT achievement test scores were the least fallible (unreliable), and the only measures achieving levels of reliability (for individual children) with which most researchers would feel comfortable.

The consequences of having individual measures of dubious reliability were considerable given the data analytic strategy employed in the longitudinal evaluation. While we may console ourselves with the belief that group means are quite reliable, it is necessary to remember that the individual child was the unit of analysis in methods employed here. Moreover, it must be remembered that these analytic methods relied heavily upon analysis of covariance procedures of various sorts, which involved correlation of individual scores to work their magic. To the extent that individual scores (as dependent or independent variables) were unreliable, covariance adjustments would not have produced the desired results. Methods for reducing the attenuating effects of unreliable measures are themselves "unreliable" for covariance designs as complicated as those required in this evaluation. But even if these methods were adequate, available estimates of individual score reliability would be inadequate. Thus, the unknown but highly questionable reliability of most child outcome measures forces us to qualify our interpretations of statistical findings.

Efforts to assess the validity of instruments in the child battery by examining intercorrelation matrices, battery factor structures, and multiple linear regressions of certain outcomes on others have not produced convincing evidence of either the validity or invalidity of the battery.[1] By this we mean, the battery has not been demonstrated either to measure or not to measure social competence, and individual measures have not been demonstrated either to measure or not to measure the traits that they were intended to measure. The efforts of previous evaluation reports to address questions of instrument and battery validity seem to have been frustrated by four major limitations. First, our understanding of the social competence construct was and is limited. Second, the scores analyzed were of far less

[1]See, for example, Interim Report VI (Granville et al., 1977).

than perfect, but unknown, reliability. Third, research of the sort necessary to seriously address questions of validity could not be conducted within the scope of the evaluation project. And fourth, the meaning of social competence, and of the PDC measures, changed dramatically from Head Start through third grade, given the nature of the world and of children in development. Thus, we encourage the reader to make his or her own judgment of the battery's face and/or content validity, and to examine correlational data presented in Appendix H for possible insight into the construct validity of child outcome measures.

One final observation is in order, the same questions that have been raised about the PDC child outcome battery might be raised about the batteries used in any large-scale educational evaluation with which we are familiar. The art of measuring human behavior, not to mention "potential," is indeed primitive. But though measurement of child outcomes in PDC was problematic, it gains strength from the diversity of measurements attempted and from a data analytic strategy that searches for patterns of effects to support interpretative leaps.

ANALYTIC DESIGN

The basic methods used to analyze child outcomes as well as other quantitative data have already been described in detail in the methods section of Chapter II--Data Analysis Methods Used in the Final Evaluation. First-order findings from Designs 1 and 2 are presented in Appendix I together with observed, predicted and ANCOVA-adjusted means for each outcome variable by group by site and by aggregate PDC and Comparison samples. A synthesis of Design 1 and 2 findings is presented in this chapter. There is only one feature of the analytic design for child outcomes that requires special explanation here--the covariate set.

Covariates Incorporated in the Analytic Design

With one exception the covariates used in analyses of child outcomes were the same as those used in analyses of parent outcomes. That exception was "impediments to parent involvement", which was included as a covariate in analyses of parent, but not child, data. Since the covariates have already been described at length in Chapter IV (Parent Outcomes), they will only be listed here:

- Sex of child
- Child's dominant language at program entry (Spanish/English)
- Child's prior (to Head Start) preschool experience (None/Some)
- Ethnicity (Anglo/Other)
- Mother's educational level (0-13+ years of schooling)

- Family structure I (two-parent structure throughout the evaluation/ one-parent structure at some point in the evaluation)
- Family structure II (stable one- or two-parent structure/unstable structure)
- Head Start test performance factor I (derived from factor analysis of average fall-to-spring Head start test scores--see Appendix H)
- Head Start test performance factor II (derived from factor analysis of average Head Start test scores--see Appendix H)

Jointly these covariates accounted for 10% to over 50% of the variance in child outcomes. Predictions were most powerful for outcomes in the domains of Specific Academic Achievement and General Academic Skill/Aptitude, followed by Learning Attitude/Style, then Attitude toward School/Teacher and Social Development/Adjustment (see Appendix I).

FINDINGS

Methods of analysis are discussed and the presentation of findings is explained in detail in Chapter II--Data Analysis Methods Used in the Final Evaluation.

Net effects of consequence are summarized graphically in Figure VI-1. All estimates of net observed and predicted effects are presented in Table VI-2; net effects judged consequential are highlighted. Site-level and aggregate findings from the synthesis of Designs 1 and 2 are presented in Tables VI-3 through VI-5 for *probable* effects, *probable and possible (?)* effects, and *probable, possible (?), and less possible (??)* effects, respectively. These tables show (at three levels of probability) the patterns of findings across time and across variables within outcome domains upon which estimates of net effect were based for each site and the total sample. Primary estimates of net effects and judgements of consequentiality were based on the patterns shown in Table VI-4 for *probable and possible* effects. Detailed results of all Design 1 and 2 analyses are reported in Appendix I.

Summary of Findings for Child Outcomes

The following research question is addressed here:

> *Did PDC children exhibit greater social competence than would have been expected had they not participated in the program?*

As we have already pointed out, measurement of "social competence" was restricted to functioning in the school setting and/or attitude toward and performance of academic tasks. Findings are briefly summarized by domain below and in Figure VI-1; they are discussed at some length in the next section.

Specific Academic Achievement

The only net effect of consequence was found at Site 3 where PDC children outperformed Comparison children on the MAT Reading Achievement Test at two points in time. There was no group difference found over all sites.

General Academic Skill/Aptitude

One net effect of consequence was found, favoring the PDC group at Site 1 and resulting from their consistently higher performance on the BSM English. No overall group difference was found.

Learning Attitude/Style

PDC-favoring net effects of consequence were found at three sites--Sites 1, 4, and 10--and over all. There were no Comparison-favoring effects.

Attitude toward Teacher/School

PDC children were found to have more positive attitudes than Comparison children at Site 4; Comparison children exhibited more positive attitudes at Site 9. There was no overall difference.

Social Development/Adjustment

The only finding of consequence was a PDC-favoring difference in Site 10. No overall difference was found.

Figure VI-1

Child Outcomes
Consequential Net Effects by Domain and Site

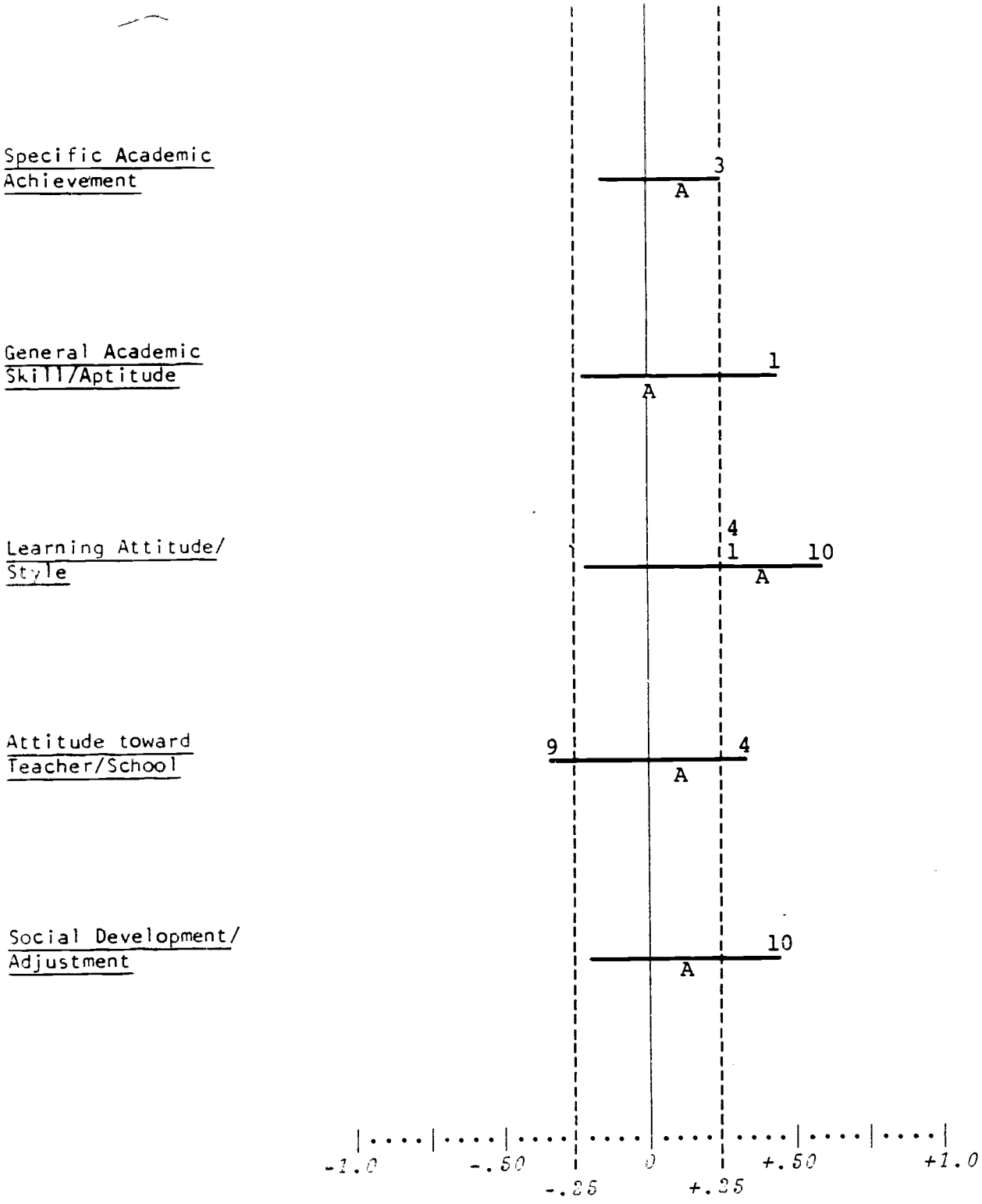


Table VI-2

Child Outcomes
Summary of Net Observed Effects and Predictions by Site
and for Aggregate PDC and Comparison Samples at Three Levels of Probability

		S1	S2	S3	S4	S5	S6	S7*	S8*	S9*	S10	Agg
Specific Academic Achievement	Probable	.13	.00	.13	.00	.00	.00	-.63	.13	.25	.00	.13
	Prob/?	.13	-.13	.25	.13	-.13	-.13	-.63	.25	.75	.00	.13
	P/?/??	.13	-.13	.25	.13	-.13	-.13	-.63	.25	.88	.29	.13
	Predicted	.00	.00	.00	.00	.00	.00	-1.0	.38	.63	-.29	.00
General Academic Skill/ Aptitude	Probable	.11	.00	.00	.22	.00	.11	.00	.00	.00	-.11	.00
	Prob/?	.44	.00	.11	-.22	.00	.22	-.22	.00	.22	-.11	.00
	P/?/??	.44	.00	.11	-.22	.00	.00	.33	-.11	.11	.00	-.11
	Predicted	-.44	.00	.00	.44	.00	.22	-.89	.22	.11	-.11	.33
Learning Attitude/ Style	Probable	.30	.00	.20	.30	-.20	-.10	-.20	.00	.00	.20	.20
	Prob/?	.30	.00	.20	.30	-.20	-.20	-.20	.10	.10	.50	.40
	P/?/??	.30	.00	.20	.30	-.20	-.20	.20	.00	.00	.60	.40
	Predicted	.00	.00	.00	-.20	.00	.10	-.70	.10	.10	-.10	.00
Attitude toward Teacher/ School	Probable	.11	.00	-.11	.11	.00	.00	.00	.00	-.22	.00	.00
	Prob/?	.00	.00	-.11	.33	-.11	.11	.11	-.11	-.33	.17	.11
	P/?/??	-.11	.00	.00	.33	-.11	.00	.11	-.11	-.33	.33	.11
	Predicted	.00	.00	-.11	-.22	.00	.22	.00	.00	.11	-.22	.00
Social Development/ Adjustment	Probable	.13	.13	-.06	.00	.06	.13	-.06	.13	.06	.19	.06
	Prob/?	.13	.19	-.19	-.19	.06	.19	.00	.06	.06	.38	.13
	P/?/??	.13	.19	-.25	-.19	.06	-.06	.38	-.13	.00	.44	.13
	Predicted	.06	.00	.19	-.06	.00	.31	-.50	.19	.06	-.06	.06

* Net observed effects for domain 1 were predicted from pretreatment characteristics and are, therefore, of dubious validity.

Table VI-3

Summary of Probable Effects for Child Outcomes
Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site
and the Aggregate Sample by Grade/Year

Outcome Domains and Measures	S1 K123	S2 K123	S3 K123	S4 K123	S5 K123	S6 K123	S7 K123	S8 K123	S9 K123	S10 K123	Agg K123
Specific Academic Achievement											
PIAT Math	0000	000+	0000	0000	0000	0000	000-	0000	00+0	0000	0000
PIAT/MAT Reading	+000	000-	+000	0000	0000	0000	----	+000	000+	000*	+000
General Academic Skill/Aptitude											
BSM English	00+0	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
Verbal Fluency	00	00	00	+0	00	0+	00	00	00	0-	00
Verbal Memory 1	0	0	0	0	0	0	0	0	0	0	0
Verbal Memory 2	0	0	0	+	0	0	0	0	0	0	0
Draw-a-Child	0	0	0	0	0	0	0	0	0	0	0
Learning Attitude/Style											
POCL 1: Task Orientation	00++	0000	0+00	0000	000-	000-	-0-0	0000	0000	0+00	0000
CI 2: Interest in Reading	00+	000	000	0++	-00	000	000	000	000	0+0	0++
CRS 3: Learning Orientation	000	000	+00	00+	000	000	000	000	000	000	000
Attitude Toward Teacher/School											
CI 1: Attitude Toward School	000	000	000	000	000	000	0+0	000	000	000	000
PI: Attitude Toward School	000	000	00-	000	000	000	0-0	000	000	0**	000
School Attendance	0+0	000	000	+00	000	000	000	000	-0-	00*	000
Social Development/Adjustment											
PIPS	0000	0++0	0000	0000	00+0	0+00	0000	0000	00+0	0000	0000
POCL 2: Sociability	00++	0000	0000	--00	0000	00+0	00-0	0000	0000	0000	0000
CRS 1: Independence	00-0	0000	000-	+000	0000	0000	0000	00+0	0000	+000	0000
CRS 2: Social Adjustment	+000	0000	0000	000+	0000	0000	0000	000+	0000	++00	+000

*Missing data.

Table VI-4

Summary of Probable and Possible (?) Effects for Child Outcomes
Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site
and the Aggregate Sample by Grade/Year

Outcome Domains and Measures	S1 K123	S2 K123	S3 K123	S4 K123	S5 K123	S6 K123	S7 K123	S8 K123	S9 K123	S10 K123	Agg K123
Specific Academic Achievement											
PIAT Math	0000	-00+	0000	0000	-000	-000	000-	++-0	+++0	0000	0000
PIAT/MAT Reading	+000	000-	+0+0	000+	0000	0000	----	+000	+0++	000*	+000
General Academic Skill/Aptitude											
BSM English	++++	0000	000+	----	0000	0000	--00	+0-0	0000	0000	0000
Verbal Fluency	00	00	00	+0	00	0+	00	00	+0	0-	00
Verbal Memory 1	0	0	0	0	0	0	0	0	+	0	0
Verbal Memory 2	0	0	0	+	0	+	0	0	0	0	0
Draw-a-Child	0	0	0	0	0	0	0	0	0	0	0
Learning Attitude/Style											
POCL 1: Task Orientation	00++	0000	0+00	+-00	000-	00+-	-0-0	0000	0000	0++0	00+0
CI 2: Interest in Reading	00+	000	000	0++	-00	00-	000	000	0+0	0++	0++
CRS 3: Learning Orientation	000	000	+00	00+	000	00-	000	+00	000	+00	+00
Attitude Toward Teacher/School											
CI 1: Attitude Toward School	000	000	000	000	000	00+	++0	+00	000	000	000
PI: Attitude Toward School	000	000	00-	000	00-	000	0-0	-0-	000	+**	00-
School Attendance	-+0	000	000	+++	000	000	000	000	---	00*	000
Social Development/Adjustment											
PIPS	0000	0+++	-0--	00-0	00+0	++00	0000	0000	00+0	0000	0000
POCL 2: Sociability	00++	0000	00+0	--0-	0000	00+0	00-0	0000	0000	00+0	00+0
CRS 1: Independence	00-0	0000	000-	+-00	0000	0000	0000	0-+0	0000	++00	0000
CRS 2: Social Adjustment	+000	0000	0000	000+	0000	0000	+000	000+	0000	+++0	+000

*Missing data.

Table VI-5

Summary of Probable, Possible (?), and Less Possible (??) Effects for Child Outcomes Derived from the Synthesis of Designs 1 and 2 and Reported for Each Site and the Aggregate Sample by Grade/Year

Outcome Domains and Measures	S1 K123	S2 K123	S3 K123	S4 K123	S5 K123	S6 K123	S7 K123	S8 K123	S9 K123	S10 K123	Agg K123
Specific Academic Achievement											
PIAT Math	0000	-00+*	0000	0000	-000	-000	+0--	++-0	+++0	0+00	0000
PIAT/MAT Reading	+000	000-	+0+0	000+	0000	0000	----	+000	++++	+00*	+000
General Academic Skill/Aptitude											
BSM English	++++	0000	000+	----	0000	-00-	--+0	+0-0	0000	000+	000-
Verbal Fluency	00	00	00	+0	00	0+	++	00	+0	0-	00
Verbal Memory 1	0	0	0	0	0	0	+	0	+	0	0
Verbal Memory 2	0	0	0	+	0	+	+	-	0	0	0
Draw-a-Child	0	0	0	0	0	0	0	0	-	0	0
Learning Attitude/Style											
POCL 1: Task Orientation	00++	0000	0+00	+-00	000-	00+-	-+-0	00-0	0000	0++0	00+0
CI 2: Interest in Reading	00+	000	000	0++	-00	00-	+++	000	0+-	+++	0++
CRS 3: Learning Orientation	000	000	+00	00+	000	00-	000	+00	000	+00	+00
Attitude Toward Teacher/School											
CI 1: Attitude Toward School	-00	000	00+	000	000	00+	++0	+00	000	000	000
PI: Attitude Toward School	000	000	00-	000	00-	000	0-0	-0-	000	+**	00-
School Attendance	-+0	000	000	+++	000	+--	000	000	---	+0*	000
Social Development/Adjustment											
PIPS	0000	0+-+	-0--	00-0	00+0	++0-	++++	00--	00+0	000+	0000
POCL 2: Sociability	00++	0000	00+0	--0-	0000	00+0	00-0	00-0	0000	00+0	00+0
CRS 1: Independence	-0-0	0000	000-	+-00	0000	0000	+00+	0-+0	0-00	++00	0000
CRS 2: Social Adjustment	++00	0000	0-00	000+	0000	--0-	+000	000+	0000	++++	+000

*Missing data.

Discussion

According to the intervention model sketched out in Chapter II, ACYF's initiative (Level A) was expected to alter institutional features (Level B) that would in turn affect parents, teachers, classrooms (Level C), and ultimately children (Level D). Though the influence of institutional factors on children would be mediated by parents and teachers, specific parent and teacher behaviors toward children were not identified in the PDC model. Institutional findings reported in Chapter III indicated that overall implementation of the basic PDC model was at best moderate and at every site somewhat inconsistent across both time and component areas. At only four sites--Sites 1, 2, 6, and 7--did implementation of the Guidelines make PDC schools/centers appreciably different overall from local Comparison institutions. The areas of greatest institutional difference between PDC and Comparison were Administration (having to do with specialized staffing and decision-making structures), Parent Involvement (involving coordination with Head Start, hiring of parents as aides, and training), and Developmental Support Services (involving coordinated provision of nutritional, medical, dental, mental health, and social services).

In evaluating parent outcomes (Chapter IV) and teacher/classroom outcomes (Chapter V) we were looking not only for impacts at that level of the intervention model (Level C) but for potential links to child outcomes. Were there outcomes for parents, teachers, and classrooms indicating that PDC children were likely to have experienced greater *developmental continuity* which the intervention model hypothesized would lead to greater *social competence*? And more practically speaking, were there specific findings for parents, teachers, and classrooms that were likely to be associated with child outcomes on the specific variables measured?

Table VI-6 pulls together all of the findings reported for all outcome domains considered in the evaluation (Chapters III-VI). Institutional outcomes suggested that local PDC programs in fact made various provisions to increase continuity. For example, generally positive findings for the Administration component of PDC indicate that specific steps were taken at most sites to involve Head Start and elementary school staff, together with parents, in joint planning of programs for children. Presumably this would have diminished any discontinuities normally experienced by children moving from Head Start into elementary school; however, we have little information about the extent of such discontinuities in participating communities, or about specific reductions in the discontinuities experienced by PDC children at Head Start-school transition. There was little evidence of effective Head Start/elementary curriculum coordination (Education Component, Subcomponent A), apparently because of school district pressures on elementary school curriculum decisions and professional differences of opinion between elementary educators and Head Start staff. On the other hand, it seems that PDC children's health records were more likely to have been transferred from Head Start to elementary school and, then, to have been consulted (by coordinators of support services). Whether this specific instance of increased continuity (health professionals

Table VI-6

Summary of Consequential Net Effects for all Outcome Domains Considered in the Evaluation by Site and for Aggregate PDC and Comparison Samples

Outcome Domain	Sites										Agg
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
INSTITUTIONAL OUTCOMES											
Administration	+	+	+	+	+	+	+		+	+	
Education				+			+?				
Bilingual/Bicultural/Multicultural		+?				+	+				
Handicapped											
Parent Involvement	+	+			+	+	+		+		
Support Services	+	+?			+?	+	+			+	
Overall Components	+?	+?				+	+				
PARENT OUTCOMES											
Involvement in School	+					-				+	
Parent as Educator	+		-				+				
TEACHER/CLASSROOM OUTCOMES											
Promotion of Parent Involvement		+	+	+	+	+			+		+
Classroom Environment		-	-					+			
Educational Management	-		-			+					-
PDC-Encouraged Instructional Content	-	+				+	-	+			+
Learning Time		-	-	-	+			-			-
CHILD OUTCOMES											
Specific Academic Achievement			+								
General Academic Skill/Aptitude	+									+	+
Learning Attitude/Style	+			+					-		
Attitude Toward Teacher/School				+						+	
Social Development/Adjustment											

in the child's environment were more likely to be aware of the child's health history and specific needs) had positive impacts on children (presumably better health) cannot be known, given available child outcome data.

Generally positive findings across sites for implementation of the Parent Involvement component of the PDC model also suggested that special efforts were made to reduce home-school discontinuity by communication with parents and by involving parents directly in the educational process. Teacher/classroom outcomes--Promotion of Parent Involvement--supported a conclusion that a number of PDC programs did substantially more to involve parents and achieved higher levels of parent involvement in the classroom. However, these findings pertained to parents in general, rather than to parents of children in the longitudinal evaluation. When we considered only outcomes for parents of children in the evaluation sample (Chapter IV), there was much less evidence of positive impact on parental involvement, and little evidence of actual (self-reported) differences in parent-child interactions related to school learning. Moreover, at only two of the three sites with PDC-favoring effects in either parent outcome domain did parent findings mesh with institutional findings (Sites 1 and 7, but not 10; see Table VI-6). Thus, parent outcomes showed little promise of linking institutional with child impacts.

The intervention model as expressed in PDC Guidelines was quite vague with respect to the ways in which teachers should behave toward children and manage the educational process to increase continuity and enhance children's social competence. The desirability of "individualizing instruction" was emphasized, but no means for accomplishing this objective was offered, much less prescribed. Rather, ACYF expected local projects to develop their own program solutions within the institutional framework provided by the general intervention model and with reference to local needs and values. Measurement of teacher/classroom outcomes (Chapter V) was intended to identify important aspects of local programs at the classroom level, assuming that distinctive PDC programs had been implemented at that level.

Findings related to teachers' behaviors toward children and children's classroom experiences (Table VI-6) did not suggest that local PDC projects developed particularly distinctive classroom-level programs. Evidence that parents were somewhat more likely to be present in PDC classrooms (a component of Promotion of Parent Involvement) at most sites meant that PDC children might have been less likely to perceive school as a world apart from their homes and communities. Findings for PDC-Encouraged Instructional Content, which included Multicultural Instruction and Use of Community Resources, also suggested that the content of children's educational experience in PDC classrooms at some sites may have been less divorced from their out-of-school experience than the normal fare served up in elementary school classrooms. In neither case, however, did our measurements indicate dramatic differences between PDC and Comparison classrooms. Furthermore, it was not at all clear how such differences might affect child outcomes in the areas actually measured.

Other findings for teachers and classrooms tended to distinguish PDC in ways that would not typically be considered desirable by most educators. In particular, net negative effects were found for Educational Management and Learning Time at a number of sites and over all sites. The PDC Guidelines did not speak to either outcome domain, and certainly did not encourage local projects to decrease the quality of management or the amount of time devoted to academic learning. But for whatever reasons (and the reasons may have been inherent in the intervention model as speculated in Chapter V) negative effects in these domains far outweighed positive indications. Thus, it seemed highly unlikely that teacher/classroom outcomes would link institutional impacts with positive impacts on children's academic learning--Specific Academic Achievement or General Academic Skill/Aptitude. In fact, if forced to establish expectations for learning outcomes, we would have predicted Comparison-favoring differences at as many as five sites (1, 2, 3, 4, and 8) as well as overall.

Regarding other child outcome domains--Learning Attitude/Style, Attitude toward Teacher/School, and Social Development/Adjustment--we had no particular expectations in the absence of strong logical or demonstrated empirical links between such child outcomes and measured outcomes for parents, teachers, and classrooms.

In sum, we approached child outcomes with very few expectations of finding PDC/Comparison differences but some suspicion that Comparison-favoring differences might emerge for learning outcomes.

We found very few differences of consequence Only two consequential net effects were observed for learning outcomes--Specific Academic Achievement and General Academic Skill/Aptitude. Both favored PDC children and both contradicted expectations based on teacher/classroom outcomes. At Site 3, where PDC teachers were judged to be significantly less successful at educational management and where children spent less time engaged in academic learning activities, a marginal (+.25) net positive effect was found for academic achievement based on differences in reading achievement in kindergarten and second grade. At Site 1, where findings were negative for Educational Management but not for Learning Time, a net positive effect was found for General Academic Skill/Aptitude based upon differences in BSM English performance from kindergarten through third grade. In neither case, have we been able to explain these findings. The reader will have already noticed in Table VI-2 that we discounted sizable net positive effects in achievement at two sites (8 and 9) and a large net negative effect at another site (7) because the observed mean differences were strongly predicted by group differences in children's pretreatment characteristics. In our judgment, achievement outcomes were clearly biased at these three sites, and we seriously doubted the power of our analytic methods to overcome the biasing effects of group nonequivalence.

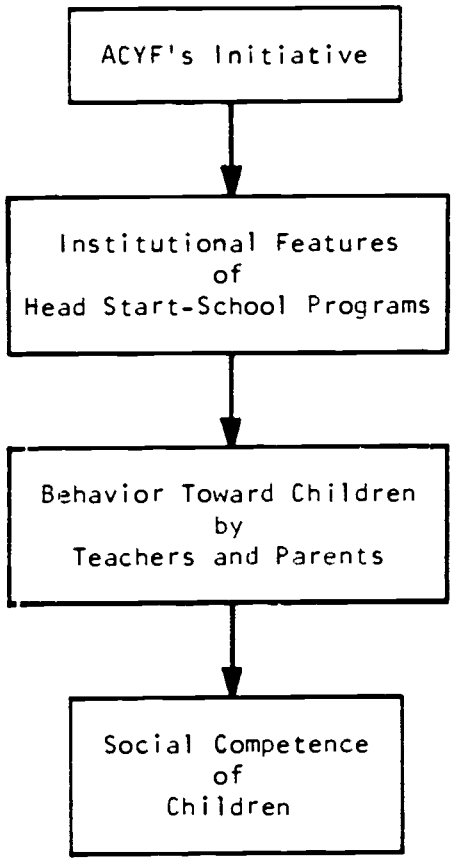
The one outcome domain in which effects were more generalized was Learning Attitude/Style where differences were particularly PDC-favoring. Net positive effects were found at three sites--Sites 1, 4, and 10--as well as over all sites. The meaning of this finding is not entirely clear since the measures employed do not have a history of use in more basic research. The POCL 1 measured children's task orientedness in test settings; ratings were made by the tester. The CI 2 measured children's interest in reading by ratings that followed an unstructured interview with each child. And CRS 3 ratings were made by class teachers and assessed the child's approach to learning and learning attitude. High scores on all three variables would suggest that a child's attitude and approach toward learning in general, and school-related learning in particular, were positive and effective, and that over the long haul such a child would function more competently both in and outside of school. Moderate (.30 to .50) correlations between these outcome measures and measures of achievement and general academic skill lend some support to this interpretation; however, PDC-favoring outcomes for Learning Attitude/Style were not paralleled by similarly positive findings for Specific Academic Achievement and General Academic Skill/Aptitude.

As a first step in determining whether parent outcomes might account for positive child outcomes, we examined the partial correlations of parent outcomes with Learning Attitude/Style variables, having removed from the latter all variance associated with the covariates used in Designs 1 and 2. The number of significant partial correlations was small and might well have occurred by chance. Moreover, these correlations were as often negative as positive. Under these circumstances, it did not seem fruitful to look any further in that direction. While it is entirely possible that some aspect of parental behavior or of classroom experience that was influenced by PDC, in turn influenced children's learning attitudes and style, the evaluation has not revealed such links.

VII

SUMMARY AND CONCLUSIONS

PDC can be viewed as a multi-step intervention, originating in ACYF's initiative (devising Guidelines, awarding grants, providing technical assistance) but requiring changes in local institutions and in the behaviors of teachers and parents to achieve its ultimate objective of enhancing children's social competence:



The evaluation of PDC assessed impacts of the intervention at each step subsequent to ACYF's initiative. Evidence of program impacts is summarized below, working backward through these steps from children--the ultimate focus of the intervention--to institutions--the initial targets of ACYF's change strategy.

FINDINGS OF PROGRAM IMPACT

Impacts on Children

There was very little evidence that local PDC programs enhanced children's social competence--the ultimate objective of ACYF's demonstration program.

At no site was children's participation in PDC associated with generally positive effects across the five outcome domains measured: Specific Academic Achievement, General Academic Skill/Aptitude, Learning Attitude/Style, Attitude Toward Teacher/School, and Social Development/Adjustment. Rather, children with Head Start backgrounds who attended PDC schools were found to be quite like Head Start graduates who attended non-PDC schools in the same communities, at least through third grade when the evaluation terminated.

The only hint of a possibly generalized PDC-favoring effect was found for the outcome domain termed *Learning Attitude/Style*. During the early elementary years, PDC children at three of ten sites were found to exhibit more positive learning attitudes/styles. Furthermore, a general PDC-favoring trend across all ten sites was indicated by aggregate tests. However, PDC-favoring findings for this domain were not paralleled by findings for other domains, and the implications of observed differences in children's learning attitude and style for later social competence are not known.

Impacts on Parents

There was little evidence that local PDC programs affected the behavior of former Head Start parents whose children were in the evaluation sample.

At only one site were PDC-favoring effects found for parents of children in the evaluation sample for both of the outcome domains measured: Involvement in School and Parent as Educator (of own child outside school). Differences in one or the other outcome domain favored PDC parents at two sites and non-PDC at two. At half of the ten sites, no differences were found in either domain.

However, parents of Head Start graduates in the evaluation sample were only a small fraction of all parents whose children attended PDC and non-PDC schools. And information about "parents in general" suggested that they were somewhat more likely to be present and actively involved with pupils in PDC than non-PDC classrooms. In fact, PDC-favoring differences were found at several sites, and a PDC-favoring trend was found over

all sites with regard to this larger group of parents. Thus, PDC parent involvement efforts may have been more successful with "parents in general" than with the small group of former Head Start parents whose children were in the evaluation sample.

There was no systematic relationship between findings for either group of parents and outcomes for children in the evaluation sample.

Impacts on Teachers and Classrooms

There was considerable evidence of difference between PDC and non-PDC teachers and classrooms; however, observed differences as often favored non-PDC as PDC teachers/classrooms and had no detectable influence on measured child outcomes.

At no site were PDC-favoring effects found for all outcome domains measured: Promotion of Parent Involvement, Classroom Environment, Educational Management, PDC-Encouraged Institutional Content, and Learning Time. However, fairly generalized cross-site effects were found within all domains except Classroom Environment. These generalized effects favored PDC teachers/classrooms in two domains and non-PDC in the other two.

Regarding Promotion of Parent Involvement, PDC teachers were more likely than non-PDC teachers to exhibit positive attitudes toward involving parents in classroom activities at two sites, and on the average over all sites; they were also somewhat more successful at actually getting parents involved at five sites, and on the average across all ten sites. These findings relate to "parents in general" rather than specifically to parents of children in the evaluation sample (see discussion of impacts on parents, above). As for the degree to which teachers emphasized PDC-Encouraged Instructional Content (health/nutrition, multicultural, community resources), site-level findings were mixed, some favoring PDC and others, non-PDC classrooms. Nevertheless, on average across all sites these aspects of curriculum tended to receive more emphasis in PDC than non-PDC classrooms. These differences between PDC and non-PDC teachers/classrooms had no obvious implications for measured child outcomes, and no relationship was found.

Differences in Educational Management and Learning Time clearly favored non-PDC over PDC teachers and classrooms. Non-PDC teachers were judged to be more effective managers of the instructional and social processes in their classrooms at two sites, and a non-PDC favoring trend was found over all sites in aggregate analyses. Non-PDC children were observed to spend more time than their PDC peers engaged in academic learning activities at four sites, and on average across all ten sites.

Observed differences in Learning Time and Educational Management favoring non-PDC clearly were not intended by ACYF but may have resulted indirectly from the PDC innovation. Specifically, secondary analyses suggested that these differences might, in part, be due to differences in the prior teaching experience of PDC and non-PDC teachers, PDC teachers being less experienced on the average. In turn, differences in level of experience appear to have been caused by selection pressures created by the PDC innovation that favored younger and less experienced teachers over older and more experienced ones. Another possible explanation, that could not be put to a quantitative test but was suggested by qualitative data, was that PDC teachers tended to allocate somewhat more time than non-PDC teachers to activities that were not strictly academic and not particularly "orderly," in keeping with PDC objectives to provide for the developmental needs of the whole child. Although these findings were worrisome, it should be noted that non-PDC-favoring differences in the amount of time devoted to academic learning by the average child were not associated with lower levels of academic skill or achievement among PDC children in the evaluation sample.

Impacts on Institutions

The institutional features prescribed by the PDC Guidelines were not fully and consistently implemented at any site, and non-PDC schools often incorporated "PDC features." Nevertheless, PDC schools were found to differ from non-PDC schools at a majority of sites in the degree to which they incorporated features associated with three components of the basic model--Administration, Parent Involvement, and Developmental Support Services.

The entire configuration of institutional features defining the basic PDC program model was not fully implemented at any site. Moreover, there were varying degrees of inconsistency over time in the implementation of particular model components at every site. Overall levels of Guideline implementation were typically moderate.

Overall differences between PDC and non-PDC schools with respect to prescribed institutional features were found at only four sites, in spite of the fact that all local PDC projects achieved at least moderate overall levels of Guideline implementation. At one site the overall similarity of PDC and non-PDC schools was clearly the result of diffusion of the PDC model within the local school district; at other sites, more complex forces at local, state, and federal levels seem to have been responsible for the institutional similarity of PDC and non-PDC schools.

The areas of greatest difference between PDC and non-PDC institutions were Administration (specialized staffing and decision-making structures), Parent Involvement (coordination of elementary school and Head Start programs, hiring of parents as aides, training of parents), and Developmental Support Services (coordinated provision of nutritional, medical, dental, mental health, and social services from Head Start through third grade). Modest PDC/non-PDC differences were found for these three components at a majority of sites.

There was little evidence of relationship between findings of institutional difference and observed impacts on teachers/classrooms, parents, and children. However, actual levels of parent involvement (for "parents in general" if not for parents of children in the evaluation sample) did tend to be higher for PDC than non-PDC samples in sites where institutional provisions for parent involvement were greater in PDC than non-PDC schools.

Summation

The findings of the evaluation suggest that (1) the program model intended by ACYF was not fully realized anywhere, (2) local versions of PDC had few impacts on parents and teachers that were likely to enhance children's social competence, and (3) children's social competence was not generally enhanced by their participation in the PDC program. Next, we consider why the evaluation failed to demonstrate PDC's effectiveness.

EXPLANATION OF FINDINGS

There would seem to be four possible explanations of why the evaluation failed to demonstrate PDC's effectiveness in enhancing children's social competence:

- The assumptions about child development underlying PDC are incorrect.
- The translation of these assumptions into action was faulty.
- The translation of these assumptions into action was occurring in participating communities independently of PDC.
- The evaluation design and methodology were inadequate to detect the program's positive effects on children's social competence.

These competing explanations are not mutually exclusive, and each may have some validity.

Incorrect Assumptions?

PDC grew out of a complex set of assumptions about factors influencing the development of social competence among children, specifically children from low income and minority backgrounds. Many of these assumptions were only implicit, and few had developed to the point of being testable hypotheses when the project got underway in 1974. The evaluation was not designed to test specific hypotheses about child development and has not done so. Rather, the evaluation was designed to determine whether a program of action shaped by this collection of assumptions would significantly improve the social competence of Head Start children during their first few years in elementary school.

In our judgment, the relationships between guiding assumptions and actual programs were so tenuous that findings of the evaluation do not significantly challenge the validity of the basic assumptions underlying PDC, nor do they lend support.

Faulty Translation?

The translation of ACYF's assumptions about factors influencing children's social competence into action involved two major steps. First, ACYF officials developed a *conceptual program* shaped by their assumptions about factors influencing the development of social competence and representing their intentions for PDC programs that would operate in field settings. Next, local projects developed *operational programs* that were supposed to realize ACYF's intentions in ways appropriate to local settings.

How well ACYF's conceptual program reflected their underlying assumptions about child development is debatable. What is not arguable is ACYF's intent that local projects would demonstrate this conceptual program in action and that we would evaluate this demonstration. Toward that end the evaluation has involved not only an assessment of program impacts but an ongoing assessment of the degree to which ACYF's intentions were actually implemented by local projects.

Findings of the evaluation raise serious questions about the fidelity of the operational programs to ACYF's intentions. Systematic evaluation of the fidelity of the operational programs was limited to those program features operationally defined in the PDC Guidelines and required of each project under the terms of their grants. Even with respect to these molar institutional features, local programs were not found to have fully or consistently implemented ACYF's intentions. If these findings are to be believed (and we believe they are), then what we evaluated were imperfect renderings of ACYF's conceptual program, truer to intentions in some aspects (e.g., provision of comprehensive nutritional, medical, dental, mental health, and social services) than in others (e.g., coordinated educational programming from Head Start through third grade).

As for how well ACYF's ultimate intentions of increasing the continuity of children's experience were realized, evidence from the evaluation is less extensive and direct. However, findings related to teacher and parent outcomes, together with the "ordinary perceptions" of site visitors, suggest that operational PDC programs did little to alter children's experience during the early years of school in ways suggested by ACYF's conceptual program. That being the case, one would not expect to find evidence of PDC's having generally enhanced children's social competence, and we did not.

In sum, available evidence strongly suggests that the translation of the PDC concept into operational programs was faulty.

The Problem of Implementation

The difficulty of implementing planned social change has received increasing attention from social scientists in recent years. All major federal demonstration projects--Follow Through, Planned Variation Head Start, and others--have experienced serious problems with program implementation, that is with getting intended programs actually implemented in field situations. PDC appears not to have been an exception to this rule, in spite of the fact that ACYF's change strategy tempered the highly directive approaches of early demonstrations with a strong reliance on *local problem solving* intended to encourage local ownership of and commitment to the program.

Volume II of this report investigates the process of program implementation at each PDC site and attempts to explain, with benefit of hindsight, why the change strategy employed by ACYF was not generally effective.

PDC Was Not Innovative?

A third explanation of why the evaluation failed to demonstrate PDC's general effectiveness might be that the PDC concept was not innovative--i.e., when implemented, did not create programs that were significantly different from other programs in the same communities. Many of the assumptions underlying PDC and the operational strategies embodied in ACYF's conceptual program were not unique to PDC, but reflected notions and values that were part of the zeitgeist. Since this same zeitgeist affected individuals and institutions at PDC sites, both indirectly and directly through other federal and state programs, it was inevitable that non-PDC schools would embody some of PDC's intended features.

State and then federal laws (P.L. 94-142) regarding the education of handicapped children gradually affected all schools considered in the evaluation, realizing many of ACYF's intentions for services to handicapped children quite independently of PDC. Concern with parent involvement was

also becoming more general within local educational systems when PDC commenced as a result of other federal programs (Title I, Title VII, Emergency School Aid Act, Follow Through) and as a result of the growing appreciation by school administrators that parental support was needed to raise school revenues during a period of declining rolls and economic retrenchment. In some measure, instructional approaches were also changing in directions intended by ACYF though quite apart from ACYF's initiative. Over the course of the project all participating school districts pressed for, and most mandated, continuous curriculum (from kindergarten, if not Head Start, through third grade and beyond), diagnostic testing, and some variety of more individualized instruction. And in one site, the PDC project, supplemented by other resources, was used to develop a bilingual program that was then diffused in large part to all schools in the community so that the district might comply with court orders.

But even though the concept of PDC became less innovative over the life of the project as a result of larger social changes, full implementation of ACYF's intentions for operational PDC programs continued to require significant change in local institutions.

Inadequate Evaluation?

It is possible, of course, that local PDC programs generally and significantly enhanced children's social competence but that the evaluation failed to detect these impacts on children. If so, the evaluation must also have failed to detect the sorts of differences between PDC and non-PDC programs, and specifically between the experiences of PDC and non-PDC children, that would have caused differences in child outcomes.

The limitations of the evaluation design and methodology are summarized briefly below:

- The power of statistical tests to detect program effects at the site level was low given small sample sizes resulting from extremely heavy attrition.

- The data analytic methods employed may have failed to control for bias in outcome measures due to observed pretreatment differences between PDC and non-PDC parents and children.

- The final analytic sample of parents and children was not fully representative of the sample entering the program in 1976, much less the larger group of parents and children served over the life of the project. And parents and children in the analytic sample may have responded less or differently to the PDC program than a more representative group would have.

- Measurement of impacts at all levels--institution, parent, teacher/classroom, and child was limited and sometimes of dubious reliability and validity. Thus, important impacts may not have been measured at all or may have been measured inadequately.

- The program's impacts, particularly on children, may only be evident after third grade, beyond the temporal scope of the evaluation.

Though we do not rule out these possible explanations of the evaluation's failure to demonstrate PDC's general effectiveness, it seems unlikely that the repeated and fairly broad-band measurements taken in this evaluation would not have revealed more evidence of impact at some level of the intervention had such impacts occurred at most, or even several, sites.

Conclusions

In spite of considerable uncertainty, we feel it our responsibility to venture a "best guess" as to meaning and implications of the evaluation's findings:

> *In general, local PDC programs do not seem to have provided children with experiences that were importantly and systematically different from the experiences of similar non-PDC children in the same communities. For this reason alone, quite apart from the evaluation's findings related to impacts on children, we do not believe that PDC generally and significantly enhanced children's social competence.*
>
> *The relative lack of significant differences in the proximate environments and experience of PDC and non-PDC children seems to have resulted from (1) incomplete and inconsistent implementation of ACYF's intentions in local PDC programs and (2) concurrent changes in non-PDC schools that significantly reduced actual and potential program differences at the local level. Problems of implementation might have been reduced by ACYF's employing a different change strategy; however, changes in non-PDC schools that made them more like the PDC model were historical "accidents" that could not have been avoided but might be viewed as partially vindicating the PDC concept.*

The evaluation of PDC was both flawed and premature. Some of its flaws could have been avoided--e.g., by having better anticipated the occurrence and limiting effects of sample attrition. Other flaws simply reflected the current state of the art--e.g., the unavailability of adequate measures of social competence and the rudimentary nature of the construct. Hindsight also suggests that the summative evaluation was undertaken prematurely, not simply before adequate methodology had been developed, but before there was any substantial evidence that local PDC programs were affecting children's experience in ways that were importantly different from what happened in the community at large.

The lessons learned from PDC do not suggest that the notion of linking Head Start with elementary school programs is either wrong or futile. Neither should we conclude that planned social change or useful evaluation is impossible. Rather, the PDC experience, together with the experiences gained in similar initiatives over the past decade, teach us humility and suggest that future efforts be more modest, focused, and informed by past experience and careful program and evaluation design.

References

Administration for Children, Youth, and Families. Project Developmental Continuity: Guidelines for an implementation year. Washington, D.C.: Author, 1975.

ACYF [Office of Child Development]. Description of the PDC program providing continuity of child development services to children and families. Washington, D.C.: U.S. Government Printing Office, 1977.

Anderson, S., & Messick, S. Social competency in young children. Developmental Psychology, 1974, 10(2), 282-293, American Psychological Association, Inc.

Bloom, B. S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bryk, A. S., & Weisberg, H. I. Use of the nonequivalent control group design when subjects are growing. Psychological Bulletin, 1977, 84(5), 950-962.

Burt, M., Dulay, H., & Hernandez, E. Bilingual Syntax Measure. New York: Harcourt, Brace, Jovanovich, 1975.

Caldwell, J. H., Huitt, W. G., & Graeber, A. O. Time spent in learning. The Elementary School Journal, May 1982, 82(4).

Campbell, D. T. Degrees of freedom and the case study. Comparative Political Studies, 1975, 8, 178-193.

Campbell, D. T., & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Disadvantaged Child: Compensatory Education a National Debate, Vol. 3. New York: Brunner/Mazel, 1970.

Carter, L. F. Federal clearance of educational evaluation instruments: procedural problems and proposed remedies. Educational Researcher, June 1977, 6(6).

Cicerelli, V. The Purdue social attitudes scales. West Lafayette, IN: Purdue University, n.d.

Cochran, W. G. The use of covariance in observational studies. Applied Statistics, 1968, 17, 270-275.

Cronbach, L. J., Rogosa, D. R., Floden, R. E., & Price, G. G. Analysis of covariance: Angel of salvation, or temptress and deluder? Stanford, CA: Stanford Evaluation Consortium, 1976.

Development Associates, Inc. & High/Scope Educational Research Foundation. A process evaluation of Project Developmental Continuity, Interim Report VIII, third program year cost report. Arlington, VA: Development Associates, Inc., 1977.

Dunn, L. M., & Markwardt, F. C., Jr. Manual for the Peabody individual achievement test. Circle Pines, MN: American Guidance Service, 1970.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliane, R., Cahen, L. S., & Dishaw, M. M. Teaching behaviors, academic learning time, and student achievement: An overview. The Journal of Classroom Interaction, Winter 1981, 17(1), 2-15.

Granville, A. C., Love, J. M., & Morris, M. A process evaluation of Project Developmental Continuity, Interim Report VII, Volume 3: Assessment of program impact through the Head Start year. Ypsilanti, MI: High/Scope Educational Research Foundation, 1977.

Granville, A. C., McNeil, J. T., Meece, J., Wacker, S., Morris, M., Shelly, M., & Love, J. M. A process evaluation of Project Developmental Continuity, Interim Report IV, Volume 1: Pilot year impact study--instrument characteristics and attrition trends. Ypsilanti, MI: High/Scope Educational Research Foundation, 1976.

Haney, W. The Follow Through planned variation experiment, Volume V: A technical history of the national Follow Through evaluation. Cambridge, MA: The Huron Institute, 1977. Prepared for the U.S.O.E.

Love, J., Granville, A., & Smith, A. A process evaluation of Project Developmental Continuity: Final report of the PDC feasibility study, 1974-1977. Ypsilanti, MI: High/Scope Educational Research Foundation, 1978.

Love, J. M., Nauta, M., Coelen, C., Grogan, M., McNeil, J., Rubin, A., Shelly, M., & Stein, M. National Home Start evaluation: Interim Report VII, twenty-month program analysis and findings. Ypsilanti, MI: High/Scope Educational Research Foundation & Abt Associates Inc.

McCarthy, D. McCarthy scales of children's abilities: Manual. New York Psychological Corporation, 1972.

McMeekin, R. W. Synthesized estimates of the costs of Head Start planned variation models. Cambridge, MA: The Huron Institute, 1973.

Morris, M., Berrueta-Clement, J., & Rosario, J. An evaluation of Project Developmental Continuity, Interim Report X: Assessment of program impact through first grade, Volume III: Impact on parents. Ypsilanti, MI: High/Scope Educational Research Foundation, 1980.

Pacific Consultants. PDC program case studies (Volumes I and II). Washington, D.C.: Author, 1979.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. Metropolitan achievement tests: Primer. New York: The Psychological Corporation, 1978a. Metropolitan achievement tests: Primary 2. New York: The Psychological Corporation, 1978b.

Rosenshine, B. V. How time is spent in elementary classrooms. The Journal of Classroom Interaction, Winter 1981, 17(1), 16-25.

Shure, M. B., & Spivack, G. The PIPS test manual. Philadelphia: Hahneman Medical College, 1974.

Smith, A., Love, J., Morris, M., Spencer, L., Ispa, J., & Rosario, J. A process evaluation of Project Developmental Continuity: Findings from the PDC implementation study. Ypsilanti, MI: High/Scope Educational Research Foundation, 1977.

Smith, M. S. Some short term effects of Project Head Start: A preliminary report on the second year of planned variation--1970-71. Cambridge, MA: Huron Institute, 1973.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. Education as experimentation: A planned variation model, Volume IV-A, an evaluation of Follow Through. Cambridge, MA: Abt Associates Inc., 1977.

U.S. Congress, House, State of the Union Message by President Lyndon B. Johnson, 90th Congress, 1st session, January 10, 1967, Congressional Record Vol. 113, pt. 1, 37.

U.S.O.E. [United States Office of Education]. The Follow Through planned variation experiment (Volumes I-V). Washington, D.C.: U.S. Government Printing Office, 1977.

Vinter, R. D., Sarri, R. C., Vorwaller, D. J., & Schaefer, W. E. Pupil behavior inventory: A manual of administration and scoring. Ann Arbor, MI: Campus Publishers, 1966.

Wacker, S., Berrueta-Clement, J., Morris, M., & Rosario, J. An evaluation of Project Developmental Continuity, Interim Report X: Assessment of program impact through first grade, Volume IV: Impact on teachers. Ypsilanti, MI: High/Scope Educational Research Foundation, 1980.

Wacker, S., Rosario, J., Diamondstone, J., & Smith, A. G. Project Developmental Continuity: Classroom observation manual. Ypsilanti, MI: High/Scope Research Foundation, 1981.

Wechsler, D. Wechsler preschool and primary scale of intelligence: Manual. New York: Psychological Corporation, 1967.

Weikart, D. P., Bond, J. T., & McNeil, J. T. The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade. Monographs of the High/Scope Educational Research Foundation, No. 3, 1978.

Wolff, M., & Stein, A. (1966). Six months later. Study I. A comparison of children who had Head Start, Summer 1965, with their classmates in kindergarten, a case study of the kindergartens in four public elementary schools, O.C.D. Project 141-61, Yeshiva University, New York.